Lexikos 31

Lexikos 31

Redakteur
Editor
Elsabé Taljard

Resensieredakteur
Review Editor
T. Harteveld

African Association for Lexicography

BURO VAN DIE WAT

STELLENBOSCH

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

Inhoud / Contents

Artikels / Articles

Projekte / Projects

Leksikohuldeblyk / Lexicotribute

Resensies / Reviews

Voorwoord

Ek is van mening dat *Lexikos* nou, te danke aan die harde werk van vorige redakteurs en bydraers, stewig gevestig is as een van die voorkeur internasionale joernale vir die publikasie van artikels oor alle leksikografiese kwessies. Een van die redes vir die gewildheid van *Lexikos* is sonder twyfel die vinnige omkeertyd, wat meegebring word deur die deurlopende aanlyn publikasieproses wat gevolg word. Ons het hierdie proses verfyn deur bladsynommers onmiddellik toe te ken by publikasie van die artikel. Outeurs hoef dus nie te wag vir die korrekte bibliografiese besonderhede van hul artikels nie. *Lexikos* is boonop in die EuroPub-indeks opgeneem.

Soos die vorige redakteur in sy voorwoord tot *Lexikos* 30 uitgewys het, het diversiteit 'n *Lexikos*-tradisie geword. Die veelsydigheid rakende die onderwerpe wat aangespreek word en die taal waarin en waaroor gepubliseer word, blyk ook duidelik uit vanjaar se *Lexikos*: hierdie uitgawe bevat bydraes van so ver as onder andere China, Montenegro, Nigerië, Pole, Slowenië en Thailand. Ek wil weer eens 'n uitnodiging aan leksikograwe van Suid-Afrika en ons buurlande in Afrika rig om hul bydraes aan ons te stuur om sodoende te verseker dat plaaslike leksikografiese uitmuntendheid die internasionale blootstelling ontvang wat dit verdien.

Die publikasie van *Lexikos* is 'n spanpoging. In hierdie verband wil ek graag me. Tanja Harteveld en me. Hermien van der Westhuizen van die WAT bedank vir hulle toewyding om te verseker dat *Lexikos* aan hul hoë tegniese standaarde voldoen. Ek wil ook graag my kollega, prof. Dion Nkomo, wat gehelp het met die keuringsproses van 'n aantal artikels wanneer die lading te veel geraak het vir my, bedank. Die redakteurskap van *Lexikos* het 'n byna voltydse taak geword, en met die steun van die AFRILEX-raad is daar besluit op 'n span redakteurs wat die werk vir elke uitgawe gelykop verdeel. Die hoofredakteurskap sal steeds roterend wees, met een van die spanlede wat die verantwoordelikheid vir die hantering van die beoordelingsproses in enige gegewe jaar aanvaar. Die hoofredakteur van *Lexikos* 32 sal my baie bekwame kollega, prof. Dion Nkomo, wees met die res van die span bestaande uit myself, dr. Steve Ndinga-Koumba-Binza en mnr. André du Plessis. 'n Spesiale woord van welkom gaan aan André, aangesien hy jong bloed en nuwe energie aan die redaksiespan verleen.

Ongelukkig het 2021 ook hartseer nuus gebring. Ons het met groot leedwese verneem van die heengaan van Sue Atkins, wat 'n groot dryfveer agter die opleiding van Afrikataal-leksikograwe was. Ons sal haar nie net as 'n belangrike leksikograaf onthou nie, maar ook as 'n deernisvolle mens, met 'n wonderlike sin vir humor. Ons het dit daarom gepas gevind om 'n huldeblyk aan Sue Atkins in hierdie uitgawe van *Lexikos* te publiseer. Prof. D.J. Prinsloo

wat haar as beide mentor en vriend geken het, het ingestem om hierdie hulde-blyk te skryf. Ons het ook verneem van die heengaan van me. Riette Ruthven, wat vanaf *Lexikos* 11 (2001) tot en met *Lexikos* 20 (2010) betrokke was by die administrasie en setwerk. Sy het altyd met entoesiasme en bedrewenheid gehelp om elke uitgawe van *Lexikos* suksesvol af te handel. Sodoende het sy 'n groot bydrae gelewer in die toenemende sukses van die tydskrif.

'n Spesiale woord van dank gaan aan ons keurders. Die keuring van artikels is 'n ondankbare en dikwels tydrowende taak, en die ontwrigting wat deur die COVID-19-pandemie veroorsaak is, het ook nie die leksikografiese gemeenskap onaangeraak gelaat nie. Die toewyding van ons keurders verseker egter dat die hoë standaard waaraan ons oor die jare gewoond geraak het, gehandhaaf word. Laastens, 'n woord van dank aan ons outeurs sonder wie se bydraes ons nie 'n tydskrif sou hê nie. Ek is dankbaar vir die positiewe gees waarmee outeurs op keurders se kommentaar reageer. Dit dra alles by tot 'n stimulerende leksikografiese gesprek.

Elsabé Taljard
Redakteur

Foreword

Thanks to the hard work of past editors and authors of contributions, I believe that *Lexikos* is now firmly established as one of the international journals of choice for publications of articles on all matters lexicographical. One of the reasons for the popularity of *Lexikos* is without a doubt the quick turn-around time, which is facilitated by the publish-as-you-go process that is followed. We have refined this process by allocating page numbers immediately on publication of the article. Authors therefore do not have to wait for correct bibliographical details of their articles. *Lexikos* has furthermore been added to the EuroPub index.

As the previous editor pointed out in his foreword to *Lexikos* 30, diversity has become a *Lexikos* tradition. The diversity with regard to topics addressed and language published in and on is also evident from this year's *Lexikos*: this edition contains contributions from as far afield as inter alia China, Montenegro, Nigeria, Poland, Slovenia and Thailand. I would once again like to extend an invitation to lexicographers from South Africa and our neighbouring countries in Africa to send us their contributions to ensure that local lexicographic excellence gets the international exposure it deserves.

The publication of *Lexikos* is a team effort. In this regard, I would like to thank Ms Tanja Harteveld and Ms Hermien van der Westhuizen of the WAT for their commitment to make sure that *Lexikos* meets their exacting technical standards. I would also like to thank my colleague, Prof. Dion Nkomo, who assisted with the adjudication process of a number of articles when the load became too heavy for me. Being editor of *Lexikos* has become an almost full-time occupation, and with the support from the AFRILEX Board it was decided to have a team of editors sharing the work equally for every edition. The main editorship will be a rotating one, with one of the team members taking the responsibility of managing the review process in any given year. The main editor for *Lexikos* 32 will be my very able colleague, Prof. Dion Nkomo, with myself, Dr Steve Ndinga-Koumba-Binza and Mr André du Plessis making up the rest of the team. A special word of welcome goes to André, since he brings young blood and new energy to the editorial team.

Unfortunately, 2021 also brought some sad news. We learnt with great sadness about the passing away of Sue Atkins, who gave impetus to the training of African language lexicographers in South Africa. We shall remember her not only as a great lexicographer, but also as a compassionate human being, with a wonderful sense of humour. We therefore saw it fit that a tribute to Sue Atkins should be published in this edition of *Lexikos*. Prof. D.J. Prinsloo who knew her both as mentor and as a friend agreed to write this tribute. We also learnt of the passing of Ms Riette Ruthven, who was involved in the administration and

typesetting processes of *Lexikos* 11 (2001) to and including *Lexikos* 20 (2010). With enthusiasm and efficiency, she assisted in the successful completion of each issue of *Lexikos*. By doing so she was instrumental in the increasing success of this journal.

A special word of thanks goes to our reviewers. Reviewing articles is a thankless and often time-consuming task, and the disruption caused by the COVID-19 pandemic has not left the lexicographic community untouched. However, the commitment of our reviewers ensures that the high standard to which we have gotten used over the years, is maintained. Finally, I would like thank our authors without whose contributions we would not have a journal. I am grateful for the positive spirit with which authors respond to reviewers' comments. It all contributes to a stimulating lexicographic discourse.

Elsabé Taljard
Editor

'n Woord van AFRILEX

Die wêreld steier steeds onder die katastrofiese gevolge van die globale COVID-19-pandemie. Hierdie globale krisis het daartoe gelei dat baie lande ongekende maatreëls soos die sluiting van skole, kolleges en universiteite in 'n omvattende grendelperiode moes toepas ten einde die volgehoue verspreiding van die nuwe koronavirus te probeer stuit. As gevolg hiervan is die mensdom uitgedaag om vorendag te kom met reaktiewe maatreëls om die ontwrigting van die lewensbestaan soos ons dit tot dusver geken het, aan te pak. Die meeste akademiese aktiwiteite het op 'n tot dusver onbeproefde en ongekende skaal na aanlyn virtuele platforms verskuif.

Na aanleiding van die ontwrigting wat die African Association for Lexicography (AFRILEX) in 2020 weens die kansellering van sy 25ste jaarlikse internasionale konferensie as gevolg van die globale pandemie ervaar het, het die AFRILEX-raad die belangrike besluit geneem om sy silwerherdenking op 'n unieke manier aan te bied. Die AFRILEX-raad het sy eerste volledig virtuele internasionale konferensie, wat goed bygewoon is deur afgevaardigdes van regoor die wêreld, suksesvol aangebied.

Die redaksie van *Lexikos* het ook daarin geslaag om deur middel van harde werk die globale storm die hoof te bied om te verseker dat hierdie vlagskip-tydskrif weer verskyn. *Lexikos* lok 'n groot aantal bydraes van plaaslike sowel as internasionale outeurs. Ten spyte van die voor-die-hand-liggende uitdagings van 'n stadige omkeertyd vir portuurbeoordeelde keuringsverslae weens die struikelblokke wat deur die globale pandemie veroorsaak is, het die redaksie van *Lexikos* nog 'n uitgawe van die tydskrif voltooi om sodoende hierdie belangrike kanaal van akademiese diens oop te hou. Daarom is besondere dank verskuldig aan die redakteur, prof. Elsabé Taljard, van die Universiteit van Pretoria, die resensieredakteur, me. Tanja Harteveld, en me. Hermien van der Westhuizen, vir hulle inspanning en volharding.

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

Namens die AFRILEX-raad, asook die algemene lede van AFRILEX, bedank ek die redaksiespan, die Buro van die WAT en die bydraende outeurs van harte vir nommer 31 van *Lexikos*. Dit is nog 'n indrukwekkende bydrae tot die kennisekonomie.

Langa Khumalo
President: AFRILEX

A Few Words from AFRILEX

The world continued to suffer the catastrophic effects of the COVID-19 global pandemic. The global crisis resulted in many countries taking unprecedented measures such as closing schools, colleges and universities in a massive lockdown in order to stem the continued spread of the novel coronavirus. As a result, humanity was challenged to come up with responsive measures to address the disruption of the life, as we hitherto knew it. Most academic activities were migrated to online virtual platforms on a hitherto untested and unprecedented scale.

Following the disruption that the African Association for Lexicography (AFRILEX) suffered in 2020, which resulted in the Association's failure to host its 25th annual international conference as a result of the global pandemic, the AFRILEX Board took a landmark decision to host its silver jubilee in a unique way this year. The AFRILEX Board successfully hosted its first fully virtual international conference, which was well attended by delegates from across the globe.

The *Lexikos* editorial team also managed to ride the global storm by working very hard to make sure that this flagship journal again sees the light of day this year. *Lexikos* continues to attract a large number of contributions from both local and international authors. Despite the obvious challenges of slow turnaround of peer-reviews owing to the complexities presented by the global pandemic, the *Lexikos* editorial team have successfully concluded another issue of this journal, thereby maintaining this important channel of academic service. To this end, I want to especially recognise the effort and tenacity of the editor, Professor Elsabé Taljard, of the University of Pretoria, the review editor, Ms Tanja Harteveld, and Ms Hermien van der Westhuizen.

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

On behalf of the AFRILEX Board, and indeed the general members of AFRILEX, I sincerely thank the editorial team, the Bureau of the WAT and the contributing authors for volume 31 of *Lexikos*. It is another telling contribution to the knowledge economy.

Langa Khumalo
President: AFRILEX

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.

Construal of Mental Health Problems in English Learners' Dictionaries

Arleta Adamska-Sałaciak, *Faculty of English,*
Adam Mickiewicz University, Poznań, Poland (arleta@amu.edu.pl)

Abstract: The purpose of the present study is to check how well monolingual learners' dictionaries cope with the sensitive field of mental ill health. The subject is considered worthy of attention in view of the high prevalence of mental health problems among young people, at whom learners' dictionaries are primarily targeted. To obtain a picture of the situation, twelve names of common mental health issues have been looked up in six learners' dictionaries of English (five British and one American). The analysis zoomed in on the choice of genus words, the information value of the definitions and examples, and the potential impact of both on the sensibilities of dictionary users. To resolve occasional disagreements in matters of content, specialist medical sources have been consulted. The treatment is cognitive linguistic in spirit, with the notion of construal — specifically its key component of focal adjustment — serving as the main descriptive tool.

Keywords: CONSTRUAL, COGNITIVE LINGUISTICS, MENTAL HEALTH, MONO-LINGUAL ENGLISH LEARNERS' DICTIONARIES, DEFINITIONS, EXAMPLES

Opsomming: Die konstruering van geestesgesondheidsprobleme in Engelse aanleerderswoordeboeke. Die doel van hierdie studie is om te bepaal hoe goed eentalige aanleerderswoordeboeke die sensitiewe veld van geestesongesteldheid hanteer. In die lig van die hoë voorkoms van geestesgesondheidsprobleme onder jongmense, op wie aanleerderswoorde-boeke primêr gerig is, verdien hierdie onderwerp aandag. Om 'n oorsig van hierdie situasie te verkry, is twaalf name van algemene geestesgesondheidskwessies in ses Engelse aanleerderswoordeboeke (vyf Brits en een Amerikaans) nageslaan. Die analise het gefokus op die keuse van verwante woorde, die inligtingswaarde van die definisies en voorbeelde, en die potensiële impak wat beide op die sensitiwiteite van woordeboekgebruikers kan hê. Om inhoudelike verskille wat soms voorkom, op te los, is spesialis- mediese bronne geraadpleeg. Die hantering is oorwegend kognitief taalkundig van aard, met die begrip van konstruering — met spesifiek die sleutelkomponent van fokusaanpas-sing — wat as die belangrikste deskriptiewe hulpmiddel dien.

Sleutelwoorde: KONSTRUERING, KOGNITIEWE LINGUISTIEK, GEESTESGESONDHEID, EENTALIGE ENGELSE AANLEERDERSWOORDEBOEKE, DEFINISIES, VOORBEELDE

1.   Rationale

Scholars have long been alert to the dangers of dealing insensitively with sensitive lexicographic material. Their attention has focused predominantly on

headwords connected with race, ethnicity, ideology, religion, gender, and age (e.g., Kachru and Kahane 1995, Murphy 1998, Swanepoel 2005, Moon 2014). Recently, Norri (2018 and 2020) added two studies relating to the field of medical practice, illness, and disability. In the present paper, I would like to look at some terms connected with mental ill health which, arguably, present an even greater challenge, if only because of the stigma they still carry in most societies.[1]

According to media reports,[2] the proportion of young people all over the world suffering from a variety of mental disorders keeps growing alarmingly. Given that monolingual English learners' dictionaries (MELDs) are available globally and are consulted predominantly by young people, it makes sense to ask how well they deal with the topic. Of course, a general-purpose dictionary is not the obvious go-to resource for mental health problems, but it is certainly one of the options.

1.1   Theoretical background

Meaning does not come ready-made, but has to be (re)constructed in the process of communication. This view, widely accepted within the humanities, has been argued for most forcefully by linguists, especially those of the cognitive persuasion. No matter which cognitive linguistic model we consult — e.g., Langacker's Cognitive Grammar, Fauconnier and Turner's Mental Spaces, Fillmore's Frame Semantics, Evans' Access Semantics — one of the pivotal ideas seems to be that individual linguistic expressions are no more than underdetermined prompts for rich meaning construction. As put by Evans (2019: 500), "linguistic expressions have meaning potential. Rather than 'encoding' meaning, [they] represent partial building instructions" from which meaning is constructed by the speaker/writer and then reconstructed by the listener/reader.

That individual linguistic expressions have meaning potentials rather than fully determined meanings is, of course, recognized also by linguistically sophisticated lexicographers (notably Hanks, e.g., 2013). Cognitive linguists, however, go a step further in that they pay special attention to how the speaker/writer 'packages' the elements to be assembled by the addressee, i.e., to what Langacker has christened construal. Resulting from our "ability to conceive and portray the same situation in alternative ways" (Langacker 2013: 445), construal is the process "whereby a linguistic expression draws upon aspects of encyclopaedic knowledge, in giving rise to a context-specific interpretation" (Evans 2019: 353). As a result, meaning is a function of both content and construal (Langacker 2000: 9). Alternative construals of the same content are always possible and no construal is ever 'correct'.

Construal focuses attention on different aspects of a scene, an operation that, by analogy with visual perception, Langacker calls focal adjustment. The adjustment is achieved by a variety of means, both lexical (e.g., the use of *came* versus *went*, or *foliage* versus *leaves*) and grammatical (e.g., the use of active ver-

sus passive, perfective versus imperfective, or a verb versus a nominalization). Focal adjustments may vary according to three parameters: selection, perspective, and abstraction (Langacker 1987). The first of these selects the domain against which a scene is construed and the entities participating in the scene. A single linguistic unit of semantic structure is typically structured by multiple conceptual domains which together form a domain matrix. In the case of the concepts investigated here, such a matrix might include, for instance, the conceptual domains of HEALTH (ILLNESS?), MIND, BEHAVIOUR, MOOD, EMOTION, or NORM. Translated into lexicographical practice, this means that a dictionary definition will situate the condition denoted by a particular headword within one of the domains (or their cross-section) by virtue of selecting the genus, such as, for example, *mental illness* or *emotional disorder*. Participants in the scene, in turn, may be associated with the choice of the differentiae, indicating which aspects of encyclopaedic (here: specialist medical) knowledge about a given condition are considered essential. Additional information can be gleaned from a dictionary's cross-references to other entries. In the online dictionaries examined, entries are frequently cross-referred to more general, thesaurus-type sections, whose names (e.g., *Mental and psychiatric disorders*) provide extra clues as to how a particular concept is construed.

The second parameter of focal adjustment, perspective, is responsible, among other things, for the distinction between subjective and objective construal. As a rule, both the speaker/writer and the addressee are construed subjectively (or offstage, to use the theatre metaphor proposed by Langacker). In objective construal, by contrast, one or both of them are placed onstage. This is achieved mainly through the use of deictics. When it comes to dictionary definitions, it is the reliance on second person pronouns (*you, your, yourself*) that construes the dictionary user as a participant in the scene, putting them onstage — an unenviable place to be in the context of mental health issues.

The last parameter, abstraction, involves the ability to conceive a situation at varying levels of schematization (Langacker 2002: 291), regulating the amount of detail given. In a dictionary definition of a medical issue, this will be realized, again, through the differentiae, thereby reflecting the lexicographer's view of how much encyclopedic information about a particular condition needs to be presented in a general-purpose dictionary.

My working assumption is that the lexicographers working on MELDs have roughly comparable knowledge regarding common mental health issues. The content of what they define will thus be more or less the same, but the construal most likely will not. Construal being largely invisible (Langacker 2000: 71), it takes some work to try and make it (more) visible. The present paper attempts to do just that, albeit on a very small scale and with reference to a very specific type of text. The belief that such an undertaking is not only feasible but also worthwhile comes, at least in part, from the steadily growing amount of work in cognitive stylistics (see, e.g., Hart 2019). The results of such stylistic

studies have demonstrated that cognitive linguistics offers tools — of which construal is but one — for analyzing all kinds of texts, not just literature.

1.2   Materials and method

Twelve lexemes from the field of poor mental health have been looked up in the free online versions of six MELDs: five British (CALD, COBUILD, LDOCE, MED, OALD) and one American (MW).[3] In order to check the validity of the author's understanding of a particular condition against state-of-the-art expert knowledge, a guidebook to the *Diagnostic and Statistical Manual of Mental Disorders* (Black and Grant 2014) has been consulted whenever neces-sary, as have recent print editions of the nine medical dictionaries listed in the References.

In addition to the umbrella term *mental illness*, most of the lexemes chosen for analysis are names of conditions, both general (*anxiety disorder, eating disor-der*) as well as more specific (*agoraphobia, anorexia, claustrophobia, bulimia, depres-sion, OCD, orthorexia*).[4] The remaining two terms (*panic attack, self-harm*) have been included on the grounds of the high prevalence of the symptoms they denote.

The main analytic focus is the entries' information content. Since allow-ances must be made for the restrictions imposed on definition writers by con-trolled defining vocabularies, less attention has been paid to the degree of defi-nitional precision than to the consistent choice of genus words, the informative value and appropriateness of examples, and the possible connotations carried by both examples and definitions (including their potential to upset, offend, or alienate the user).

2.   Analysis

2.1   Quantitative

As is evident from Table 1, two dictionaries, CALD and MED, feature all the entries we are interested in. Seven of the twelve entries (*depression, obsessive-compulsive disorder/OCD, agoraphobia, claustrophobia, eating disorder, anorexia, bulimia*) are present in all six dictionaries, one (*self-harm*) is included in five, another one (*orthorexia*) in four, and three (*mental illness, anxiety disorder, panic attack*) in two dictionaries. This is not a bad result, especially when compared with the coverage of the same terms in medical dictionaries, as summarized in Table 2.[5] None of the latter has *orthorexia* (which, incidentally, is absent also from DSM-5) and only one features *self-harm* (which DSM-5 lists as a newly proposed disorder under the category *Conditions for Further Study*). Thus, when it comes to phenomena which have entered public discourse relatively recently,[6] MELDs actually do better than medical dictionaries. This is hardly

surprising. After all, before being able to include any given term in a specialist dictionary, its authors have to wait for the relevant official body to approve the term, together with its associated concept, as part of the discourse of their discipline.

Table 1:   Coverage of entries denoting mental health issues in MELDs

	mental illness	depression	anxiety disorder	OCD (or full form)	agora-phobia	claustro-phobia	panic attack	eating disorder	anorexia (nervosa)	bulimia	orthorexia	self-harm
CALD	+	+	+	+	+	+	+	+	+	+	+	+
COBUILD	-	+	-	+	+	+	-	+	+	+	-	+
LDOCE	-	+	-	+	+	+	-	+	+	+	+	+
MED	+	+	generalized a.d.	+	+	+	+	+	+	+	+	+
OALD	-	+	-	+	+	+	-	+	+	+	+	+
MW	-	+	-	+	+	+	-	+	+	+	-	-

Table 2:   Coverage of entries denoting mental health issues in medical dictionaries

	mental illness	depression	anxiety disorder	OCD (or full form)	agora-phobia	claustro-phobia	panic attack	eating disorder	anorexia (nervosa)	bulimia	orthorexia	self-harm
Black's	+	+	a. neurosis/ state	+	+	+	+	+	+	+	-	-
Collins	-	+	*under* anxiety	+	+	+	+	-	+	+	-	-
DoMT	+	+	+	+	+	+	+	+	+	+	-	+
Dorland's	m. disorder	+	+	+	+	+	p. disorder	+	+	+	-	-
Melloni's	+	+	-	+	+	+	-	-	+	+	-	-
Mosby's	+	+	+	+	+	+	+	+	+	+	-	-
Stedman's	+	+	+	+	+	+	+	+	+	+	-	-
Taber's	*under* illness	+	+	+	+	+	p. disorder	+	+	+	-	-
Webster's	+	+	+	+	+	-	+	-	+	+	-	-

2.2   Qualitative

MENTAL ILLNESS

As might be expected, none of the MELDs has an entry for *mental health issue*, *mental health problem* or *mental disorder*, so *mental illness* seemed the obvious

place to start. CALD and MED, the only two dictionaries that feature the entry,[7] offer paraphrases ('an illness that affects the/someone's mind'). In LDOCE, a definition of *mental illness* can be gleaned from the entry for *mentally ill*: 'having an illness that affects your mind and your behaviour'. The mention of behaviour adds an important element to the definition, but the use of the second person pronoun, placing the reader onstage, is unfortunate. As we shall see, this is not the only headword in which objective construal is encountered.

DEPRESSION

How depression is viewed differs from one dictionary to the next. In addition to the expected ILLNESS, we also find the superordinate domains of STATE and CONDITION. The actual genus words are: *mental illness* (CALD), *mental state* (COBUILD), and *medical condition* (LDOCE, MED, MW, OALD). *Mental illness*, although not technically wrong, might be questioned on the grounds that it is a strong, possibly stigmatizing, description, so an argument could perhaps be made for reserving it for psychotic disorders. On the other hand, the somewhat vague *mental state* is not necessarily better, since it might be seen as downplaying the seriousness of depression. On balance, *medical condition*, especially when modified by *serious* (MW), seems the best description, not least because it carries an indirect promise that medicine can help the sufferer. This is not to say that it would satisfy everyone: many therapists would no doubt object to *medical condition* on the grounds that it implicitly privileges the biochemical aetiology of depression. However, such an objection would amount to hairsplitting if made with regard to a general-purpose dictionary.

Depression is the most polysemous item among those considered here. The uses we are interested in, that is, those pertaining to mental/emotional states, fall into two related categories: the medical one mentioned in the preceding paragraph versus the more everyday one — a deep feeling of sadness. Most dictionaries reflect that. CALD, MW and OALD offer two distinct senses: in CALD and MW the medical condition comes second, in OALD it is defined first. LDOCE and MED both present the two readings of *depression* as subsenses within one sense, but order them differently: LDOCE prioritizes the medical condition, while MED starts with the 'feeling of being extremely unhappy'. COBUILD alone does not acknowledge that *depression* may be used in two senses, defining it simply as '... a mental state in which you are sad and feel that you cannot enjoy anything, because your situation is so difficult and unpleasant'. This does not convey the idea of illness at all, nor does it allow for the possibility of endogenous depression, where the depressive state is somatic in origin rather than caused by external stressors (cf. *Melloni's*).

CALD, LDOCE, OALD, and MED, the dictionaries that do a reasonably good job of defining depression in the medical sense, offer some degree of detail (i.e., exhibit a low degree of abstraction) by specifying how a depressed person feels (CALD and LDOCE: 'very unhappy and anxious', MED: 'unhappy',

OALD: 'very sad and anxious'). The CALD, LDOCE and OALD definitions thus seem marginally superior to the MED one in that they refer to anxiety, a frequent symptom of depression. Additionally, CALD, LDOCE and MED note that a depressed person is unable to lead a normal life, while OALD mentions 'physical symptoms such as being unable to sleep'.

COBUILD and LDOCE resort to objective construal. The COBUILD case has been quoted above; LDOCE says: 'a medical condition that makes you very unhappy and anxious and often prevents you from living a normal life'. It would have been just as easy to use *someone* or *a person* instead of *you*.

The examples on offer are mostly useful, often providing some insight into the social context of depression, e.g., CALD: *Tiredness, loss of appetite, and sleeping problems are all classic symptoms of depression.*, *Withdrawal is a classic symptom of depression.*; MED: *Some children show signs of anxiety and depression at exam time.*; OALD: *She suffered from severe depression after losing her job.*; MW: *Many people suffer from clinical depression for years before being diagnosed.* LDOCE gives one example at the relevant sense (featuring *post-natal depression*) and seven more corpus examples, four of which can stand on their own.[8]

ANXIETY DISORDER

Only CALD features the headword in this form. MED, the only other dictionary that covers the condition, redirects automatically to *generalized anxiety disorder*. The genus in both definitions is *mental illness*, one 'in which a person is so anxious that their normal life is affected' (CALD) or 'that causes someone to feel very anxious most of the time' (MED). MED's selection of differentiae is perhaps more suggestive of what it feels like to suffer from anxiety. It should also be noted that CALD includes NORM in the conceptual domain matrix. Reference to norms of any kind is always risky, since it implies abnormality, or at least non-conformity with the norm. In this case, however, it is not normality as such that is invoked, but a person's 'normal life', so the definition's potential for being interpreted as judgemental is relatively low.

The single example given in CALD (*Panic attacks, irrational fears, and compulsive behaviour are examples of anxiety disorders.*) is perfect. Mentioning compulsions as one of the manifestations of pathological anxiety is especially useful, since non-specialists may well be unaware of the link between the two. MED's example (*People with symptoms of generalized anxiety disorder tend to always expect disaster.*) also has some added value relative to the definition.

OBSESSIVE-COMPULSIVE DISORDER (OCD)

Two entries have been examined here: that for the full name of the condition and for the commonly used abbreviation. LDOCE and MW do not have *obsessive-compulsive disorder*, but both have entries for *OCD* (where the abbreviation is expanded) and for the adjective *obsessive-compulsive*. The respective defini-

tions in LDOCE are markedly different: the one for *OCD* concentrates on the symptoms ('a person does the same thing again and again and cannot stop doing it'), while the one for the adjective goes some way towards explaining what lies behind a sufferer's behaviour ('they have strong anxious feelings'). MW has no definition under *OCD*; its definition of *obsessive-compulsive* features the genus *mental illness,* similarly to CALD, LDOCE, and MED. OALD opts for *mental disorder*. MW's differentiae consist of symptoms ('repeating actions or thinking about certain things too much'). It is also the only dictionary that explicitly mentions obsessive thoughts; the others limit themselves to compulsive actions. MED (which has *obsessive-compulsive disorder* and cross-refers *OCD* to it) and OALD (which does the opposite) both provide explanations of compulsive behaviour in terms of goals: 'in order to avoid painful thoughts' (MED) and 'to get rid of fears or unpleasant thoughts'(OALD). COBUILD restricts itself to behavioural rituals ('... cannot stop doing a particular thing, for example washing their hands'). CALD is the least helpful: its 'mental illness that causes a person to do something repeatedly for no reason' is at once simplistic and vague. No concrete symptoms are mentioned at all (by contrast, COBUILD, LDOCE, and MED all speak of hand-washing), nor is there any attempt at explaining why sufferers act the way they do. All dictionaries except MED ('a mental illness that makes you keep repeating an action…') manage to avoid objective construal.

Except for LDOCE's 'strong anxious feelings', no explicit connection is made anywhere between OCD and anxiety. This is regrettable, since most medical sources (*Dorland's*, *Melloni's*, *Mosby's*, *Stedman's*, *Taber's*, *Webster's*) treat OCD as a type of anxiety disorder. On the whole, after consulting medical dictionaries, the general impression is that MELDs tend to underplay the seriousness of the condition. *Melloni's* and *Dorland's* say that compulsions and obsessions 'interfere with personal or social functioning'; *Mosby's* that they 'significantly interfere with the patient's occupational, social, or interpersonal functioning'; *Taber's* that they 'interfere with effective living'. According to *Collins Dictionary of Medicine*, obsessive thoughts 'are so frequent and intrusive as to cause distress or disability' and '[t]here is often depression and sometimes suicide'. The popular stereotype of a compulsion, i.e., repetitive hand-washing, the only example on offer in some MELDs, may also unwittingly trivialize the condition. Perhaps the higher level of schematization in those dictionaries that do not mention any typical compulsive actions at all is a better solution.

Only LDOCE has full-sentence examples. Three of the four of them add valuable information to the definition (two are about the co-occurrence of OCD and anorexia, the third one mentions Prozac as an approved treatment).

AGORAPHOBIA

A different conceptual domain (FEAR) is selected here than for the headwords analyzed so far. Of course, the fact that the genus word in all six dictionaries is

fear may result from purely etymological considerations, but it may also indicate that the referent is construed as a symptom rather than an illness in its own right. The latter interpretation would agree with *Black's*, according to which 'agoraphobia is a *symptom* of psychological disorder' [emphasis added], but not, e.g., with *Mosby's*, where it is treated in the same way as OCD, that is, as a type of anxiety disorder.

Five of the learners' dictionaries define *agoraphobia* as the fear of open or public spaces or places; only CALD transcends the etymology. After mentioning the basic 'fear of going outside and being in public places', it goes on: 'or of being in a situation from which it might be difficult to escape or in which help might not be available'. This is in line with current psychiatric wisdom. According to DSM-5 (Black and Grant 2014), agoraphobia may involve anxiety in response to finding oneself in an enclosed place, such as a theatre or a train. This means that agoraphobics sometimes present with symptoms traditionally associated with claustrophobia, which is also remarked upon in the *Collins Dictionary of Medicine*, albeit only in the *claustrophobia* entry. To the uninitiated, the comorbidity of agoraphobia and claustrophobia might seem paradoxical since, in popular understanding, *agoraphobia* and *claustrophobia* are antonyms, a fact reflected in the MELDs' cross-references (only COBUILD does not make the connection).

CALD gives two examples, one of which (*Suffering from agoraphobia, she's afraid to even step outside her home.*) appears to be constructed rather than corpus-based, but nonetheless adds detail to the definition. LDOCE has seven sentences in its *Examples from the corpus* section, two of which seem particularly informative: one accentuates the counterintuitive relationship between agoraphobia and claustrophobia mentioned above (*It was agoraphobia but felt like claustrophobia.*), while the other offers a clue as to the origins of agoraphobia, linking it with another common symptom (*When some one can't cope with panic attacks, agoraphobia is often the consequence and treating this condition requires specialist help.*)[9] This agrees with *Stedman's*, which says that agoraphobia is 'often associated with panic attacks'.

CLAUSTROPHOBIA

The domain selected is, again, predominantly FEAR, except in COBUILD, whose full-sentence definition ('Someone who suffers from claustrophobia feels very uncomfortable or anxious when they are in small or enclosed places.') requires no genus, and thus no commitment to a conceptual domain. This contrasts with the same dictionary's treatment of *agoraphobia* ('Agoraphobia is the fear of open or public places'.) — an obvious inconsistency.

Apart from FEAR, another domain, that of FEELING, makes a brief appearance in two definitions. OALD lumps the psychiatric sense of 'an extreme fear of being in a small confined place' with the extended — possibly figurative — sense of 'the unpleasant feeling that a person gets in a situation which restricts

them', separating the two by a semicolon. MW does likewise, except that 'an unhappy or uncomfortable feeling caused by being in a situation that limits or restricts you' is here given the status of a separate second sense. The objective construal, signalled by the use of *you* above, could easily have been avoided, as demonstrated by the definitions in the remaining dictionaries.

CALD's only example reads: *He suffers from claustrophobia so he never travels on underground trains.* The majority of LDOCE's (eight) corpus examples do not go beyond illustrating common collocations (*suffer from claustrophobia, a feeling of claustrophobia*). One is the same as in the *agoraphobia* entry (*It was agoraphobia but felt like claustrophobia*); another one (*To be here, to have this happening, the claustrophobia of their fatuous intrigues?*) is context-dependent to the point of being meaningless.

PANIC ATTACK

Only CALD and MED have an entry for *panic attack*. CALD's differentiae cover more symptoms ('your heart beats fast, you have trouble breathing, and you feel as if something very bad is going to happen') and thus provide more information. However, the genus word chosen (*period*) seems odd, and the whole phrase ('a sudden period of severe anxiety') may be difficult to process. MED's choice of domain (FEELING) is, arguably, better, but the definition, while shorter and simpler ('a sudden very strong feeling of being afraid or worried that makes you unable to breathe'), ends up being slightly vague. One also wonders whether *worried* is not too weak a description of the feeling of impending doom which is typically part and parcel of a panic attack (see, e.g., *Stedman's*).

More information pertaining to domain selection can be gleaned from the thesaurus links: CALD cross-refers to the sections *Mental and psychiatric disorders* and *Anxiety and worry — general words*, while MED has links to *Fear and fright* and to *Feelings of worry and nervousness*. Whether this is enough to help the average user make the connection between, on the one hand, panic attacks and, on the other, anxiety disorder and agoraphobia is less than certain.

However, the main problem in both definitions is objective construal. The use *you* and *your*, which places the dictionary user firmly onstage, could easily have been avoided, as *someone* or *a person* would have worked in both cases.[10]

Neither dictionary offers examples of use.

EATING DISORDER

As in the case of depression, there is no agreement among the dictionaries regarding the domain against which eating disorders are best viewed. Two selections (ILLNESS and CONDITION) are repeated, and a new one, DISORDER, appears as well, most likely because it echoes the form of the headword. CALD has the genus *mental illness*; COBUILD, LDOCE, and MED all have *medical condition*; OALD plumps for *emotional disorder.* The choice of genus in CALD,

LDOCE, and MED is thus consistent with the treatment of the term *depression* in those dictionaries.

When it comes to differentiae, the domain of NORM is referenced everywhere except in CALD. OALD talks about 'eating habits that are not normal', COBUILD about not eating 'in a normal or healthy way'. LDOCE invokes normality in a subtler way, focusing on the amount of food consumed, but, unfortunately, at the same time resorting to objective construal ('you do not eat a normal amount of food'). The other dictionaries use *someone* (MED), *person* (COBUILD), *people* (CALD), or manage not to make reference to a human subject at all (OALD). On the whole, CALD's definition ('a mental illness in which people eat far too little or far too much food and are unhappy with their bodies') appears to be both the most informative and the least likely to cause discomfort.

COBUILD and MED feature both anorexia and bulimia in their definitions; OALD mentions anorexia; LDOCE has a cross-reference to the *bulimia* entry.

COBUILD offers one example (*Anyone can develop an eating disorder, but young women are most vulnerable.*). CALD has three examples 'from literature', one of which (*Eating disorders mean eating compulsively because of a distorted body image.*) reinforces the information from the definition on the possible aetiology of the disorder. One of the three examples in LDOCE (*Thus the compulsive relationship with physical exercise can become effectively part of the spectrum of eating disorder behaviour.*) adds important information about eating disorder symptomatology (compulsive exercising) that is not mentioned in any of the definitions.

ANOREXIA (NERVOSA)

One would expect the domain selected here to be the same as for the superordinate category of eating disorders or, alternatively, that the genus would be *eating disorder* itself. However, only CALD and OALD are consistent, the former sticking with *mental illness*, the latter with *emotional disorder*. COBUILD, LDOCE, MED, and MW agree with CALD that anorexia is an illness (specified as *mental* in LDOCE and *physical and emotional* in MW). It may be a moot point whether anorexia is less serious than, say, depression (which, as we remember, OALD defines as a *medical condition* rather than as an *emotional disorder*); what is not in doubt is that it is notoriously difficult to treat and all too often has a fatal outcome.[11] That being the case, further qualification of the illness as *serious* (in CALD, MED, and MW) is fully justified.

On this occasion, only one dictionary resorts to objective construal (MED: 'a serious illness that makes you want to stop eating').

COBUILD, MW, and OALD point to the fear of becoming fat as the motivation behind anorexic behaviour. None of the dictionaries mentions compulsive exercising (cf., e.g., *Taber's*) as one of the symptoms.

CALD, MW, and OALD issue warnings with regard to the effects of ano-rexia (CALD: 'often resulting in dangerous weight loss'; MW and OALD: 'leading to dangerous weight loss'). The same cannot be said of COBUILD. Making no overt mention of the danger to sufferers' health and emphasizing the thinness resulting from anorexia ('a person has an overwhelming fear of becoming fat, and so refuses to eat enough and becomes thinner and thinner'), the definition might even be (mis)interpreted as a veiled encouragement to food avoidance.

MED and OALD say that anorexia affects mainly young women, a fact highlighted in MW's only example (*Anorexia most commonly affects teenage girls and young women*). CALD also gives a single example; LDOCE has seven sen-tences from the corpus, most of them either too context-dependent or too reliant on difficult lexical items (*demise, incidence, prevalence*) to be of much use.

CALD, MED, MW, and OALD have cross-references to *bulimia*; addition-ally OALD cross-refers to *orthorexia*. OALD also has links to thesaurus sections on *Healthy eating habits* and *Unhealthy eating habits*, from where the inquisitive user can arrive at the general domain of HEALTH, with subdomains such as DIET, FITNESS, ILLNESS, MEDICINE, and MENTAL HEATH, a cross-section that gives a pretty good idea of where the anorexia concept is situated.

BULIMIA

Analogically to anorexia, CALD, COBUILD, LDOCE, MED, and MW all place bulimia within the general domain of ILLNESS (in this instance qualified as *mental* only in CALD and as *physical and emotional* in MW), while OALD sticks with *emotional disorder*. MED and MW, justifiably, describe bulimia symmetri-cally to anorexia, i.e., as *serious*. Unlike with anorexia, where both MED and OALD state that the illness affects mainly young women, here it is MW that makes the connection. One cannot help but feel that this is indicative of a lack of co-ordination in the treatment of the two headwords. Another inconsistency is the absence of objective construal in MED — proof, if proof were needed, that it could have been avoided also in the case of anorexia.

When it comes to the level of abstraction, all the dictionaries mention vomiting after eating, and all except COBUILD and MED refer to the large quantities of food consumed. Additionally, CALD and LDOCE highlight the sufferer's lack of control over their food consumption (CALD: 'someone eats in an uncontrolled way'; LDOCE: 'a person cannot stop themselves from eating too much'). Three dictionaries offer some insight into the aetiology of bulimia (COBUILD: ' a person has a very great fear of becoming fat'; MED '... in order to control his or her weight'; MW: '... in order to not gain weight').

Four of the dictionaries have no examples at all. CALD's only example 'from literature' (*This is bulimia, the eating disorder discussed earlier.*) is distinctly unhelpful. LDOCE gives six sentences from the corpus. Two of those (*Unlike anorexia nervosa, bulimia survives by disguise., Often women with bulimia repeat past*

patterns of behaviour in current relationships.) are potentially informative, though both may be too cryptic to be of help to the average user.

CALD, MED, and OALD have cross-references to anorexia. Additionally, CALD has a link to *Diets and dieting* and to *Mental and psychiatric disorders*, and OALD to *Healthy eating habits* and *Unhealthy eating habits*.

ORTHOREXIA (NERVOSA)

Two hitherto unencountered domains are invoked here: OBSESSION and DESIRE. This must be because orthorexia is not (yet?) unanimously recognized as a medical problem, as evident from its absence from medical dictionaries. CALD defines it as an *obsession*, MED as an *extreme obsession*, and LDOCE as an *extreme desire*. It cannot go unremarked that while *desire* is neutral, *obsession* has negative connotations. However, choosing the latter genus is not necessarily a bad thing if one believes that orthorexia poses a danger to a person's physical and mental well-being. CALD, agreeing with MED and LDOCE about the intensity of the condition, in fact introduces the domain of mental illness in the rest of the definition: 'so extreme it can be considered a mental illness'. OALD — rather surprisingly in view of its take on *anorexia* and *bulimia* (both defined as *emotional disorders*) — opts for *medical condition*.

Only OALD features an example sentence, one that could serve as a self-sufficient definition (*A person with orthorexia will be obsessed with defining and maintaining the perfect diet*). The same dictionary also has cross-references to *anorexia* and *clean eating.*

SELF-HARM

There is no such entry in MW. The remaining dictionaries feature *self-harm* either as a verb (COBUILD, OALD), a noun (LDOCE, MED), or both (CALD). Whatever the actual reason behind the choice of grammatical category, it seems that deciding on the verb leads to problems. Verbs encode different meanings than nouns or nominalizations due to differences in construal: the former rely on sequential scanning, the latter on summary scanning (Langacker 2002: 78). Nominalization involves a cognitive operation of reification (Talmy 2006: 78) which, apart from the obvious — construing the referent as an entity rather than a process — imposes more distance between, on the one hand, the speaker/writer and the addressee and, on the other, the situation described. The upshot, in the lexicographic context, is that when defining a verb, there is a danger of putting the dictionary user onstage, which is exactly what happens in three dictionaries,

According to CALD and COBUILD, the verb means 'deliberately hurt yourself', according to OALD, 'deliberately injure yourself'. As with the definitions of *panic attack* and *eating disorder* criticized above, here, too, the use of *yourself* is gratuitously direct. Things look worse still when one reads on:

'because of mental illness' (COBUILD), 'as a result of having serious emotional or mental problems' (OALD), 'because you have emotional problems or are mentally ill' (CALD). This picture of the dictionary user as a likely self-harmer would have been easiest to avoid in COBUILD, had its full-sentence definition started with 'If someone/a person self-harms ...' rather than 'If you self-harm ...'. In CALD and OALD, where a reflexive pronoun had to be used, *oneself* would have done the trick. Even assuming that their style-guides require CALD and OALD lexicographers to use *yourself*, surely cases such as this one warrant a departure from the rule.

It may have been precisely in order to steer clear from *yourself* that LDOCE and MED chose to present *self-harm* as a noun. Their definitions are near-identical, the common part reading: 'physical harm that someone (…) deliberately does to their own body, for example (by) cutting their skin'. MED post-modifies the *someone* ('with emotional problems'), while LDOCE graphically elaborates on how they hurt themselves ('with a knife'). In fact, all the dictionaries mention cutting, which results in a relatively low level of abstraction.

CALD gives a morphosemantic definition (see Geeraerts 2003: 90) of the noun *self-harm* as *act of self-harming*. This might be questioned on the grounds that it is the individual act that is profiled here rather than its recurrent nature.

There are no example sentences in LDOCE. COBUILD and MED have one each, with *she* as the subject; OALD has one, with *I* in the subject position (*As a teenager I was self-harming regularly*). CALD gives two examples, one where the self-harmer is a *he*, the other, more informative one, talking about *people* (*The number of people admitted to hospitals after deliberate self-harm has increased over the last 20 years.*).

CALD cross-refers to *Mental and psychiatric disorders*, MED, rather less insightfully, to *Injuries and wounds*.

3.   Concluding remarks

An attempt has been made in this study to employ the notion of construal, in particular its component of focal adjustment, as an ancillary tool for the analysis of dictionary entries. Some differences have been found in the way individual MELDs construe mental health issues, though it is far from certain that they always stem from a conscious decision on the lexicographers' part.

Let us take domain selection first. CALD is the most consistent in its choice of genus, with *depression, obsessive-compulsive disorder, anxiety disorder, eating disorder, anorexia,* and *bulimia* all gathered under the umbrella of *mental illness* (and with *anorexia* additionally qualified as *serious*). This seems to genuinely reflect the way the specific conditions are construed. One might perhaps wonder whether the consistency is not achieved at the cost of precision, especially when we compare CALD with OALD, which distinguishes between *emotional disorder* (for *eating disorder, anorexia,* and *bulimia*), *mental disorder* (for *OCD*) and *medical condition* (for *depression* and *orthorexia*). However, the OALD cate-

gories need not be the result of carefully considered choices. The reason behind opting for *disorder* may simply have been that this lexical element is present in both *eating disorder* and in the full form of *OCD*. Why the OALD genus for *orthorexia* is the same as for *depression* and not, as one would expect, the same as for *anorexia* or *eating disorder*, remains a mystery. On balance, then, the superficially finer distinctions in OALD do not necessarily indicate differences in construal.

When it comes to perspective, objective construal appears at least once in the definitions of the following eight headwords: *mental illness* (LDOCE), *depression* (COBUILD, LDOCE), *obsessive-compulsive disorder* (MED), *claustrophobia* (LDOCE, MW), *panic attack* (CALD, MED), *eating disorder* (LDOCE), *anorexia* (MED), *self-harm* (CALD, COBUILD, OALD). In other words, no dictionary manages completely to avoid placing the dictionary user onstage, although, to be fair, MW and OALD have each erred only once. Importantly, a lot of this seems haphazard: why should *you* appear in the *claustrophobia* entry but not in the *agoraphobia* one, why in *anorexia* but not in *bulimia*? The easily achievable switch to subjective construal is perhaps the most important practical recommendation following from the present study. Since dictionaries alone do not have the power to eradicate the stigma associated with mental health issues, they should at least avoid implying that the user is, or may be, a sufferer. And yes, a dictionary 'talking to the user' is a commendable project, but the recommendation does not need to be mechanically followed in every single entry.

As regards the level of abstraction, we have seen that the amount of detail given in the definitions varies. When it comes to examples, LDOCE has the most to offer, with CALD's 'examples from literature' coming a close second. A number of the examples, in all the analyzed dictionaries, add little or nothing to the definitions. In the case of names of mental health problems, as with encyclopaedic entries in general, the semantic function of examples is at least as important as their grammatical or stylistic function. Perhaps lexicographers should set themselves more ambitious goals than illustrating usage by listing common collocations (e.g., *suffer from X*) or quoting unedited corpus examples, some of which have zero informative value when taken out of context.

I have approached the topic of this study with the conviction that *mental illness*, while appropriate as a hyperonym for severe disorders such as psychosis, is best avoided, on the grounds of delicacy, in the case of the conditions discussed here. I was therefore hoping that medical dictionaries might suggest some more suitable, less explicit alternatives for domain selection. That hope remained largely unfulfilled. With some of the conditions, it is the behavioural aspect that comes to the fore, for instance, *Mosby's* defines *eating disorders* as a *group of behaviours*. This is fine when the definition is long and exhaustive, supplying a lot of further detail, but would be less so in the necessarily concise treatment in a general-purpose dictionary. For *depression*, we get *mood disorder* (*Taber's*) and *abnormal emotional state* (*Mosby's*). While *mood disorder* seems spot on, it might be too difficult for the less advanced users of learners' dictionaries. The

other option, *abnormal emotional state*, would be more acceptable without the NORM domain being invoked, but then *emotional state* alone is too vague. *Taber's* genus *social phobia* for *agoraphobia* is, again, exactly right in terms of classification, but of debatable value for a learners' dictionary.

Finally, the reader might expect to be told, on the basis of the above comparison, which dictionary comes out on top. While there is no clear all-round winner, it is CALD that comes closest, definitionwise, to what we might expect from a responsible reference work, attuned to the cognitive and emotional needs of its users. CALD features all the examined entries and exhibits the greatest consistency in domain selection. Except in the case of *obsessive-compulsive disorder*, it also gives the fullest, most informative definitions.

In sum, appreciating the numerous difficulties and restrictions that MELD writers have to grapple with, and the sheer impossibility of making sure that every semantic field is dealt with as a system of related concepts, I would nonetheless argue that the domain of mental health is too important, especially given the predominantly young audience of learners' dictionaries, to not be treated with special care. Raising mental health awareness may not be the primary goal of pedagogical lexicography, but if it can be achieved with relatively minor modifications to current lexicographic practice — consistent choice of genus terms, sticking with subjective construal, choosing informative examples and editing them where needed — it would be a shame not to pursue it.

Endnotes

1.   The 'still' may be an oversimplification; see Burek Pierce (2010) for an account of how depression (in English) was firmly connected to physical health, and thus managed to avoid stigmatization, until as late as the 17th century. For more background on the interaction between our perception of disease and the language used to talk about it, see Sontag (1978).

2.   See https://www.theguardian.com/society/2018/nov/22/why-do-more-young-people-have-mental-health-problems.

3.   Full dictionary titles and other bibliographical details are given in the References.

4.   The centrality of depression, OCD, phobic and eating disorders among contemporary mental health issues has been confirmed time and again by psychopathologists (see, e.g., Ugazio 2013).

5.   If there is no main entry for a particular term, but the medical dictionary cross-references it to another entry, the entry is treated as present. This seems to be the most sensible solution in view of the partly onomasiological organization of specialist dictionaries.

6.   The earliest OED3 citation for *orthorexia* comes from 1997.

7.   Here, as in many other cases, COBUILD redirects to CED. While it is easy for the user to follow the links, the decision has been made not to include CED in the analysis, since, as a dictionary addressed to native speakers, it is not subject to similar defining restrictions as the other reference works looked at here. Besides, I would like the analysis to be valid also for the print editions of the dictionaries in question.

8.   *Mild symptoms of anxiety and depression are often associated with social difficulties. She suffers from periods of deep depression, when she locks herself away and will speak to no one for weeks. He has been*

suffering from depression since his wife died last year. During the past few decades, prescription drugs have also been widely used to control the symptoms of depression.

9. Cf. *Webster's New World Medical Dictionary*: 'Persons with agoraphobia frequently also have panic disorder'.

10. Incidentally, that would also have been in accordance with the practical precepts formulated for COBUILD definitions by Hanks (1987: 126).

11. See, e.g., the *Collins Dictionary of Medicine*.

References

A.   Dictionaries

General-purpose

CED = Collins Online English Dictionary. Glasgow: HarperCollins. Date of access: 1 April 2019.
    https://www.collinsdictionary.com/dictionary/english

CALD = Cambridge Advanced Learner's Dictionary and Thesaurus. Cambridge: Cambridge University Press. Date of access: 1 April 2019.
    https://dictionary.cambridge.org/pl/dictionary/english/

COBUILD = Collins COBUILD Advanced English Dictionary. Glasgow: HarperCollins. Date of access: 1 April 2019.
    https://www.collinsdictionary.com/dictionary/english/

LDOCE = Longman Dictionary of Contemporary English. London: Pearson. Date of access: 1 April 2019.
    https://www.ldoceonline.com/

MED = Macmillan English Dictionary. London: Macmillan Education. Date of access: 1 April 2019.
    https://www.macmillandictionary.com/

MW = Merriam-Webster's Advanced Learner's English Dictionary. Springfield, Massachusetts: Merriam-Webster, Incorporated. Date of access: 31 August 2019.
    http://learnersdictionary.com/

OALD = Oxford Advanced Learner's Dictionary. Oxford: Oxford University Press. Date of access: 1 April 2019.
    https://www.oxfordlearnersdictionaries.com/

OED3 = Oxford English Dictionary online. Date of access: 22 January 2020.
    https://www-1oed-1com-10186786x12c9.han.amu.edu.pl/

Medical

Black's Medical Dictionary. 2010. 42nd edition. London: A&C Black.

Collins Dictionary of Medicine. 2007. 4th edition. Glasgow: Harper Collins.

Dictionary of Medical Terms. 2004. 4th edition. London: A&C Black.

Dorland's Illustrated Medical Dictionary. 2012. 32nd edition. Philadelphia: Elsevier Saunders.

Melloni's Illustrated Medical Dictionary. 2010. 4th edition. New York/London: Informa Healthcare.

Mosby's Medical Dictionary. 2013. 9th edition. St Louis: Elsevier Mosby.

Stedman's Medical Dictionary. 2006. 28th edition. Baltimore: Lippincott Williams & Wilkins.

Taber's® Cyclopedic Medical Dictionary. 2017. 23rd edition. Philadelphia: F.A. Davis.

Webster's New World™ Medical Dictionary. 2008. 3rd edition. Hoboken, NJ: Wiley Publishing.

B.   Other references

Black, D.W. and J.E. Grant. 2014. *DSM-5® Guidebook: The Essential Companion to the Diagnostic and Statistical Manual of Mental Disorders.* 5th edition. Washington, DC: American Psychiatric Publishing.

Burek Pierce, J. 2010. Defining Health: Melancholy and Mutation in Early Modern and Modern Medical Dictionaries. Adams, M. (Ed.). *"Cunning Passages, Contrived Corridors": Unexpected Essays in the History of Lexicography*: 147-163. Monza, Italy: Polimetrica International Scientific Publisher.

Evans, V. 2019. *Cognitive Linguistics: A Complete Guide*. Edinburgh: Edinburgh University Press.

Geeraerts, D. 2003. Meaning and Definition. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 83-93. Amsterdam/Philadelphia: John Benjamins.

Hanks, P. 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 116-136. London: HarperCollins.

Hanks, P. 2013. *Lexical Analysis*. Cambridge, MA: The MIT Press.

Hart, C. (Ed.). 2019. *Cognitive Linguistic Approaches to Text and Discourse: From Poetics to Politics.* Edinburgh: Edinburgh University Press.

Kachru, B.B. and H. Kahane (Eds.). 2013 [1995]. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta.* Berlin: Mouton de Gruyter.

Langacker, R. 1987. *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, R. 2000. *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter.

Langacker, R. 2002. *Concept, Image, and Symbol*. Berlin/New York: Mouton de Gruyter.

Langacker, R. 2013. *Essentials of Cognitive Grammar*. Oxford: Oxford University Press.

Moon, R. 2014. Meanings, Ideologies and Learners' Dictionaries. Abel, A. et al. (Eds.). *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15–19, 2014*: 85-105. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.

Murphy, M.L. 1998. Defining People: Race and Ethnicity in South African English Dictionaries. *International Journal of Lexicography* 11(1): 1-33.

Norri, J. 2018. Definitions of Some Sensitive Medical Words in Dictionaries of English. *International Journal of Lexicography* 31(3): 253-273.

Norri, J. 2020. Treatment of Words for Illness and Disability in Monolingual English Dictionaries. *International Journal of Lexicography* 33(3): 227-250.

Sontag, S. 1978. *Illness as Metaphor.* https://monoskop.org/images/4/4a/Susan_Sontag_Illness_As_Metaphor_1978.pdf.

Swanepoel, P. 2005. On Defining "Imaginary" Beings and Attributes: How Do Lexicographers Cope with Culturally Determined Differences in Beliefs about Cosmology, Ontology and Epistemology? *Lexikos* 15: 179-195.

Talmy, L. 2006 [1988]. The Relation of Grammar to Cognition. Geeraerts, D. (Ed.). 2006. *Cognitive Linguistics: Basic Readings:* 69-108. Berlin/New York: Mouton de Gruyter.

Ugazio, V. 2013. *Semantic Polarities and Psychopathologies in the Family: Permitted and Forbidden Stories.* Translated by Richard Dixon. New York/London: Routledge.

Zur Darstellung eines mehrstufigen Prototypbegriffs in der multilingualen automatischen Sprachgenerierung: vom Korpus über *word embeddings* bis hin zum automatischen Wörterbuch

María José Domínguez Vázquez, *Universidade de Santiago de Compostela — ILG, Spain (majo.dominguez@usc.es)*

Zusammenfassung: Das multilinguale Wörterbuch zur Substantivvalenz *Portlex* gilt als Ausgangpunkt für die Entstehung der automatischen Sprachgeneratoren *Xera* und *Combinatoria,* deren Entwicklung und Handhabung hier präsentiert wird. Beide Prototypen dienen zur automatischen Generierung von Nominalphrasen mit ihren mono- und biargumentalen Valenzstellen, die u.a. als Wörterbuchbeispiele oder als integrierte Bestandteile künftiger autonomer E-Learning-Tools eine Anwendung finden könnten. Als Modelle für neuartige automatische Valenzwörterbücher mit Benutzerinteraktion fassen wir die Sprachgeneratoren in ihrem heutigen Zustand auf.

Bei dem spezifischen methodologischen Verfahren zur Entwicklung der Sprachgeneratoren stellt sich die syntaktisch-semantische Beschreibung der vom Valenzträger eröffneten Leerstellen aus syntagmatischer und paradigmatischer Sicht als Schwerpunkt heraus. Zusammen mit Faktoren wie der Repräsentativität, der grammatischen Korrektheit, der semantischen Kohärenz, der Frequenz und der Vielfältigkeit der lexikalischen Kandidaten sowie der semantischen Klassen und der Argumentstrukturen, die feste Bestandteile beider Ressourcen sind, sticht ein mehrschichtiger Prototypsbegriff hervor. Die kombinierte Anwendung dieses Prototypbegriffs sowie von *word embeddings* zeigt zusammen mit Techniken aus dem Gebiet der maschinellen Verarbeitung und Generation natürlicher Sprache (NLP und NLG) einen neuen Weg zur künftigen Entwicklung von automatisch generierten plurilingualen Valenzwörterbüchern.

Insgesamt stellt der Beitrag die Sprachgeneratoren sowohl aus der Perspektive ihrer Entwicklung als auch aus Nutzersicht dar. Der Fokuss wird auf die Rolle des Prototypbegriffs bei der Entwicklung der Ressourcen gelegt.

Schlüsselwörter: NLG: *NATURAL LANGUAGE GENERATION*, AUTOMATISCHES WÖRTERBUCH, INTERAKTIVES WÖRTERBUCH, SPRACHGENERATOREN, KORPUSLEXIKOGRAPHIE, ONTOLOGIE, PROTOTYP, LEXIKALISCHER PROTOTYP, SEMANTISCHE PROTOTYPISCHE KLASSEN

Abstract: Towards the Description of a Multi-sided Prototype Concept in Multilingual Automatic Language Generation: From Corpus via *Word*

Embeddings **to the Automatic Dictionary.** The multilingual dictionary of noun valency *Portlex* is considered to be the trigger for the creation of the automatic language generators *Xera* and *Combinatoria*, whose development and use is presented in this paper. Both prototypes are used for the automatic generation of nominal phrases with their mono- and bi-argumental valence slots, which could be used, among others, as dictionary examples or as integrated components of future autonomous E-Learning-Tools. As samples for new types of automatic valency dictionaries including user interaction, we consider the language generators as we know them today.

In the specific methodological procedure for the development of the language generators, the syntactic-semantic description of the noun slots turns out to be the main focus from a syntagmatic and paradigmatic point of view. Along with factors such as representativeness, grammatical correctness, semantic coherence, frequency and the variety of lexical candidates, as well as semantic classes and argument structures, which are fixed components of both resources, a concept of a multi-sided prototype stands out. The combined application of this prototype concept as well as of *word embeddings* together with techniques from the field of automatic natural language processing and generation (NLP and NLG) opens up a new way for the future development of automatically generated plurilingual valency dictionaries.

All things considered, the paper depicts the language generators both from the point of view of their development as well as from that of the users. The focus lies on the role of the prototype concept within the development of the resources.

Keywords: NLG: *NATURAL LANGUAGE GENERATION*, AUTOMATIC DICTIONARY, INTERACTIVE DICTIONARY, LANGUAGE GENERATORS, CORPUS LEXICOGRAPHY, ONTOLOGY, PROTOTYPE, LEXICAL PROTOTYPE, SEMANTIC PROTOTYPICAL CLASSES

1.   Einführung

Der Ausgangspunkt für die Gestaltung der Sprachgeneratoren *Xera* und *Combinatoria* (Domínguez Vázquez et al. 2020) ist das online Wörterbuch zur Substantivvalenz *Portlex* (Domínguez Vázquez et al. 2018). *Portlex* ist ein semikollaboratives, multilinguales, *cross-lingual* Wörterbuch für Deutsch, Spanisch, Französisch, Italienisch und Galicisch (Domínguez Vázquez und Valcárcel Riveiro 2020). Bei seiner Entwicklung haben wir feststellen können, dass die Erhebung von Korporabelegen[1] zur Veranschaulichung der vom Substantiv eröffneten Valenzstellen sowie insbesondere zur Beschreibung des Kombinationspotentials der verschiedenen Ausdrucksformen jedes Arguments für fünf Sprachen viel Zeit in Anspruch nimmt. Besonders bei Suchanfragen zur Verdeutlichung der Interaktion der Ergänzungen hat sich eine nicht zu unterschätzende Anzahl an erhobenen Korporabelegen angesichts des Wörterbuchtyps nicht immer als adäquat erwiesen. Das betrifft vor allem die Realisierung der Ergänzungen in Form eines Adjektivs oder eines Kompositabestandteils, denn — wie allgemein bekannt — geht die Korpushäufigkeit einer Ausdrucksform nicht unbedingt mit ihrem Ergänzungsstatus einher.

Aufgrund der Tatsache, dass die verwendeten Korpora zur Anfertigung von *Portlex* weder funktional-syntaktisch noch semantisch annotiert sind[2] —

dementsprechend lassen sich die Korporabelege und -daten nicht nach diesen Parametern abrufen —, haben wir uns dazu entschieden, die Perspektive umzukehren: anstatt Korporabelege zur Darstellung der Nominalvalenz anzubieten, korpusgestützte Beispiele für die jeweiligen Valenzmuster automatisch zu generieren. So sind *Xera* und *Combinatoria* entstanden.

Beide Prototypen vermitteln automatisch generierte Nominalphrasen für jeweils 20 Substantive im Spanischen, Deutschen und Französischen[3]. *Xera* erzeugt einfache syntaktisch-semantische Argumentstrukturen — wie z.B. *der Geruch nach Blumen*; *Combinatoria* trägt zur Vermittlung komplexerer nominaler Valenzmuster bei, wie z.B *der Geruch des Zimmers nach Blumen* (siehe 3.). Ferner sind die Generatoren zur Einbindung weiterer Beschreibungseinheiten sowie Sprachen gestaltet und lassen sich ebenfalls in weitere Ressourcen integrieren und dadurch wieder verwerten.

In beiden Ressourcen ist eine valenzfundierte Beschreibung des nominalen Kombinationspotentials mit Fokus auf die kombinatorische Bedeutung (Engel 1996, 2004) von unentbehrlicher Bedeutung. Es geht hierbei nicht nur um die Frage, ob bei der Wiedergabe einer semantischen Rolle eine bestimmte ontologische Entität eine Valenzstelle belegen kann (oder nicht), sondern auch darum, welche konkreten lexikalischen Kandidaten bzw. ontologischen Kategorien diese Stelle einnehmen können. Diese Herangehensweise hängt mit dem Hauptziel von Valenzwörterbüchern zusammen, die Schumacher (2006: 1396) so beschreibt:

> Ein Defizit dieser [grammatikographischen] Darstellungsform besteht darin, dass hierbei die Regeln immer nur mit relativ wenigen Beispielen illustriert werden können. Dadurch bleibt für den Benutzer die Frage bestehen, welche der Regeln für die vielen Fälle zutreffen könnten, die nicht als grammatikalisches Beispiel dienen. Hier kann ein Lexikon Abhilfe schaffen [...].

Bei der Anfertigung der Sprachgeneratoren hat sich die methodologische Frage, wie man eine semantischfundierte Beschreibung der nominalen Valenzstellen automatisch erlangt und generiert, als eine Herausforderung erwiesen. Zu ihrer Beantwortung sind wir von der Anwendung einer „kombinierten Methode" mit Rückgriff auf die Interoperabilität von Ressourcen ausgegangen. Im Konkreten baut das methodologische Verfahren zur Festlegung der anzubietenden nominalen Argumentstrukturen sowie der Beispiele auf (a) einer valenzfundierten Beschreibung des Substantivs und der zu belegenden Leerstellen, (b) einer korpusgestützten Analyse erhobener Ko-Okurrenzen sowie ihrem Kombinationspotential, (c) eine automatische Datenerhebung von lexikalischen Daten aus NLP-Ressourcen und aus den Wordnets für die beschriebenen Sprachen, vorhanden in dem *Multilingual Central Repository*[4] (s. Domínguez Vázquez, Solla Portela und Valcárcel Riveiro 2019) und (d) einem mehrschichtigen Prototypbegriff auf.

Prototypen sind in unserem Ansatz typische bzw. repräsentative Instanzen für eine konkrete Slotbesetzung oder für eine konkrete Argumentstruktur. Der Beschreibung des Prototypbegriffs weist dieser Beitrag einen gesonderten

Stellenwert zu, denn er gilt als Verknüpfungselement bei der automatischen Datenerhebung und -generierung. Weiterhin sind die von Generatoren gelieferten Beispiele und Muster im Hinblick auf ihre Repräsentativität, semantische Kompabilität sowie grammatische Korrektheit, Vielfältigkeit und Prototypizität sowohl manuell als auch automatisch herausgefiltert worden[5]:

— Es ist allgemein bekannt, dass grammatische Korrektheit mit semantischer Akzeptabilität nicht immer einhergeht, sowie dass eine Ausdrucksform mit unterschiedlichen Bedeutungen korrelieren kann. Daher wird die syntaktische und semantische Analyse der vom Valenzträger geforderten Argumenten, ihrer Interaktions- sowie (In)kompatibilitätsregeln und ihrer Distribution unter Berücksichtigung valenzfundierter Kriterien sowie der Korrektheit und Akzeptabilität zum Schlüsselkonzept. Es handelt sich hier um die Analyse der kombinatorischen Bedeutung im Sinne von Engel (2004: 188).

— Zur Gewährleistung repräsentativer und prototypischer Beispiele werden zum einen Belege aus den Korpora aus *Sketch Engine* herangezogen[6] und nach Häufigkeitskriterien sortiert, zum anderen bedarf die Entfaltung des im Nomen enthaltenen Potentials einer näheren Untersuchung, die mit Rückgriff auf den Prototypbegriff sowie auf die Anwendung der prädiktiven Methode *word2vec* (*word embeddings*) erfolgt.

— Die semantische Zuordung des Sprachmaterials und die Etablierung von semantischen Klassen wird mithilfe der Ontologien aus WordNet[7] erlangt. Dazu greift man ebenfalls auf die Prototypizität zurück.

Der Fokus des Beitrags liegt somit auf der Darstellung des methodologischen Verfahrens und des den Generatoren zugrundeliegenden mehrschichtigen Prototypbegriffs, d.h. auf der Entwicklung der Generatoren. Das ist u.E. insofern relevant, dass sie ein computergestütztes Modell für die Erstellung von automatisch generierten plurilingualen Valenzwörterbüchern vorstellen. Einblicke in die Nutzerperspektive lassen sich auch gewinnen.

Der Beitrag ist wie folgt aufgebaut: Im Abschnitt 2 wird ein Gesamtüberblick über Grundeigenschaften der Generatoren gegeben. Abschnitt 3 verdeutlicht angesichts eines mehrschichtigen Prototypbegriffs das methodologische Verfahren. Auf die Wechselwirkung zwischen Typizität, Frequenz und Repräsentativität bei der Auswahl der Argumentstrukturen, lexikalischer Kandidaten und semantischer Klassen wird in 4 eingegangen. Der 5. Abschnitt setzt sich mit den Generatoren als Modell für ein automatisches plurilinguales Wörterbuch auseinander.

2.   Allgemeines zur Typologie der multilingualen Sprachgeneratoren

Das wissenschaftliche Interesse an der natürlichsprachlichen Generierung (NLG) ist seit den 90er Jahren gewachsen. Heutzutage gibt es unterschiedliche auto-

matische Generatoren natürlicher Sprache, deren Ziel die Erzeugung menschen-ähnlicher Texte jeglicher Art ist. Als Beispiele lassen sich meteorologische oder sportliche Berichte, medizinische Zusammenfassungen, vereinfachte Texte, Empfehlungstexte und Dialoge, u.a. (Vicente et al. 2015, Nallapati et al. 2016, Sordoni et al. 2015) anführen. Ebenfalls ist es möglich, aus Bildern Texte automatisch zu generieren und umgekehrt (Otter et al. 2020) sowie mit einer geringen Menge an Ausgangsmaterial Witze, Gedichte und Geschichten (Roemmele 2016) zu erstellen. Auf lexikographischem Gebiet rechnet man auch mit Vorschlägen zur automatischen Generierung von Wörterbüchern (Bardanca Outeiriño 2020, Héja und Takács 2012, Kabashi 2018) sowie Wörterbucheinträgen (Geyken et al. 2017), zur Identifizierung von mikrostrukturellen Teilen, wie z.B. den Beispielen (Kilgarriff et al. 2008), oder zu ihrer automatischen Erhebung (Kosem et al. 2019). Das Endprodukt unserer Generatoren ist ein anderes, und zwar syntaktisch-semantisch akzeptable einfache und komplexe Nominalphrasen im Kontext[8]. In beiden Generatoren handelt es sich nämlich um die Vermittlung einer ausreichenden und distinktiven Auskunft über den vom Valenzträger festgelegten syntaktisch-semantischen Rahmen, d.h. über die syntaktisch-semantische Schnittstelle der Nominalphrase.

3.   Allgemeine Beschreibung der Sprachgeneratoren *Xera* und *Combinatoria*

3.1   Beschreibungsebenen: die Festlegung der Argumentstruktur

Jeweils 20 Substantive des Deutschen, Spanischen und Französischen, deren Auswahl auf ihre Einordnung in unterschiedliche Szenen zurückgeht, sind zur Auswertung der Prototypen herangezogen worden (Tab. 1).

Szenen	Substantive
BEWEGUNG	huida ǀ Flucht ǀ fuite viaje ǀ Reise ǀ voyage mudanza ǀ Umzug ǀ déménagement
LOKATION	presencia ǀ Anwesenheit ǀ présence ausencia ǀ Abwesenheit ǀ absence estancia ǀ Aufenthalt ǀ séjour
AUSDRUCK	conversación ǀ Gespräch ǀ conversation discusión ǀ Diskussion ǀ discussion pregunta ǀ Frage ǀ question respuesta ǀ Antwort ǀ réponse texto ǀText ǀ texte video ǀ Video ǀ vidéo

AFFIZIERTHEIT	muerte ǀ Tod ǀ mort
	aumento ǀ Zunahme ǀaugmentation
	dolor ǀ Schmerz ǀ douleur
	amor ǀ Liebe ǀ amour
KLASSIFIKATION	olor ǀ Geruch ǀ odeur
	sabor ǀ Geschmack ǀ saveur
	color ǀ Farbe ǀ couleur
	anchura ǀ Breite ǀ largeur

Tab. 1:   Substantive in *Xera* und *Combinatoria*

Für die Gewährleistung der automatischen Erhebung und Generierung sprachlicher Daten sind unterschiedliche Arbeitsphasen und -verfahren notwendig: (i) Festlegung der Argumentstruktur, (ii) semantische Prototypisierung, (iii) Expandierung der Prototypen, iv) Debugging und v) paradigmatische Verpackung und automatische Generierung der Nominalphrase (s. Domínguez Vázquez, Solla Portela und Valcárcel Riveiro 2019). Die unterschiedlichen Arbeitsschritte und die dafür eingesetzten Ressourcen und Tools fasst die Abb. 1[9] zusammen:
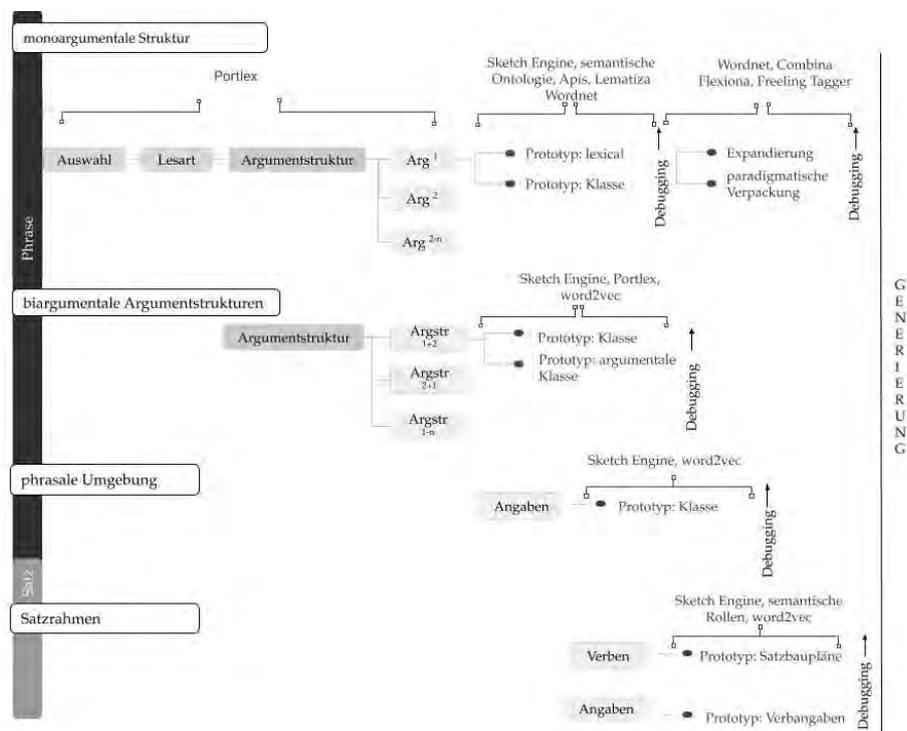

Abb. 1:   Beschreibungsebenen und Ressourcen

Die erste Beschreibungsphase[10] betrifft die Festlegung der Argumentstruktur, was am Beispiel des Substantivs TEXT veranschaulicht wird: Abb. 2[11] zeigt die Analyse der Valenzstellen, Abb. 3 einige biargumentale Muster.
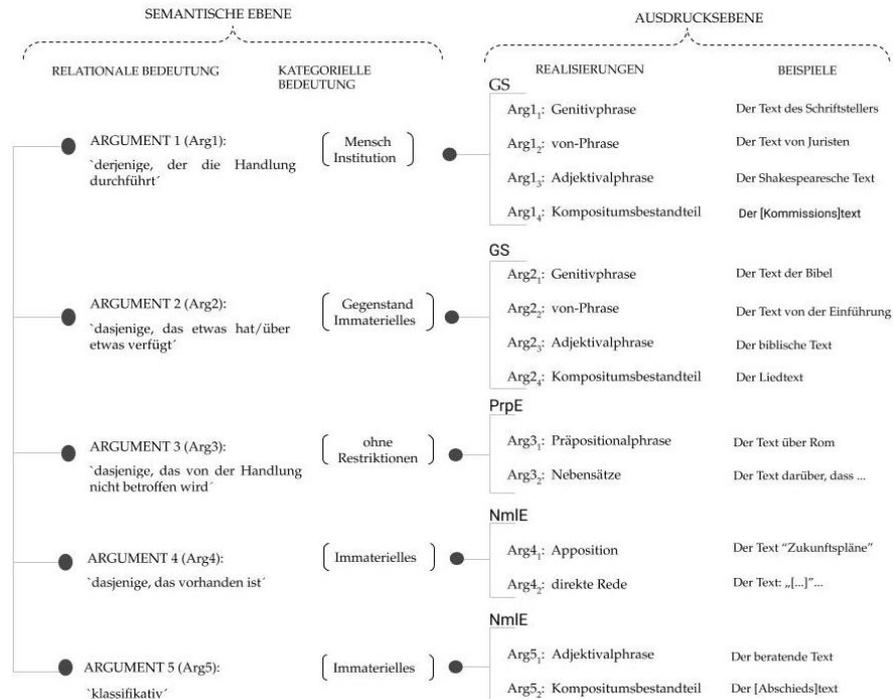
Abb. 2:   Argumentstellen von TEXT

Abb. 3:   Biargumentale Muster von TEXT

Wie aus der Abb. 2. hervorgeht, steht in unserem Modell ein semantisch verankerter valenzfundierter Ansatz im Vordergrund:

— Die jeweiligen Argumentstellen werden nach ihrer relationalen Bedeutung — wie bei Arg1 `derjenige, der die Handlung durchführt´ — analysiert. Die Beschreibung der semantischen Rollen ermöglicht den notwendigen Unterschied zwischen einem Arg1$_4$, z.B. 'Kommissionstext', gegenüber einem Arg5$_2$ wie 'Abschiedstext'.

— Die Kandidaten zur Wiedergabe einer Rolle bzw. zur Besetzung einer bestimmten Leerstelle bestimmen wir ebenfalls kategoriell[12] (s. 4.3.) und schreiben sie einer semantischen Klasse (s. 4.4.) zu.

Das bereits Angeführte sollte nicht den Eindruck erwecken, dass die Ausdrucksebene außer Acht gelassen wird, ganz im Gegenteil: eine vollständige Darstellung valenzbezogener Argumentstrukturen setzt die Beschreibung möglicher Realisierungen sowie ihres Kombinationspotentials voraus, und dies für die drei beschriebenen Sprachen.

3.2   Nutzer und Nutzung der Ressourcen

FremdsprachenstudentInnen und -dozentInnen stellen sich als den an erster Stelle anvisierten Benutzerkreis heraus. Beide können bei unterschiedlichen Benutzungszielen und -situationen von den Ressourcen profitieren, vor allem wenn man in Betracht zieht, dass (i) die auf die Valenz zurückgehenden Fehler im Fremdsprachenunterricht häufig vorkommen (Gao und Liu 2020, Nied Curcio 2014) und (ii) das Erlernen des Wortschatzes eng mit der Aneignung der syntaktisch-semantischen Umgebung zusammenhängt (Laufer und Nation 2012). Als sekundäre Nutzer kommen Wissenschaftler und Computerlexika, die sprachliches Wissen benötigen, in Frage (vgl. 5).

Bei der Handhabung beider Tools ist vom menschlichen Nutzer zunächst die Sprache und das nachzuschlagende Substantiv auszuwählen. Anschließend entsteht in beiden Tools ein Dropdown-Menü: das Argumentstrukturmenü in *Xera* (Abb. 4) und eine zu selegierende Kombination von Argumentenstellen in *Combinatoria* (Abb. 7).
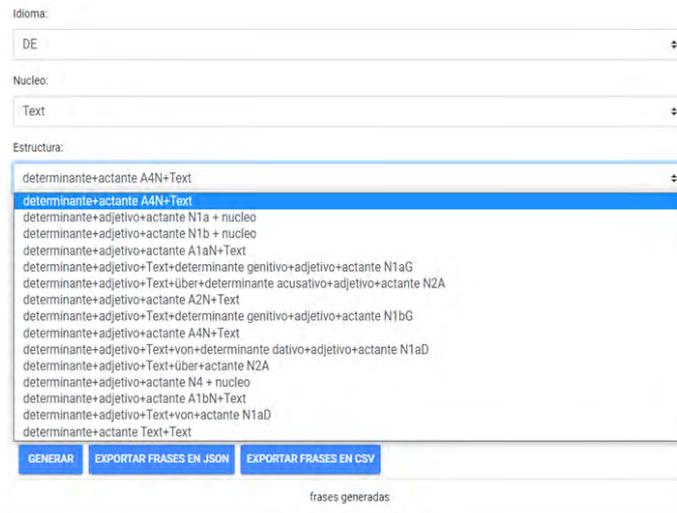
Abb. 4:   Argumentstrukturmenü bei *Xera*. Beispiel TEXT

Je nach selegiertem Muster wird ein weiteres Menü mit den semantischen pro-
totypischen Klassen (s. 4.4.) der ausgewählten Leerstelle entfaltet. Somit liefert
Xera beim Muster [Determinant + Adjektiv + TEXT + Determinant im Genitiv +
Arg1[1]] die folgenden semantischen Klassen[13], begleitet von Standardbeispielen:
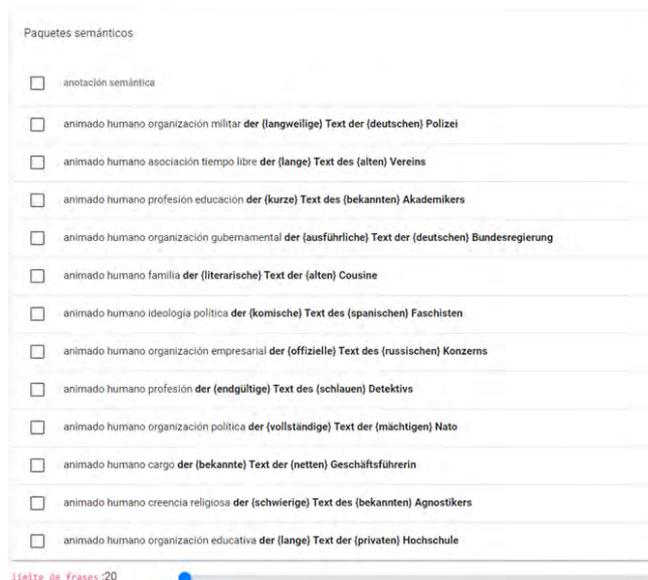
Abb. 5:   Auswahlmenü semantischer Klassen bei *Xera*. Beispiel TEXT

Selegiert man in *Xera* dann eine konkrete semantische Klasse, z.B. {ANIMADO HUMANO PROFESIÓN EDUCACIÓN}[14], werden Beispiele zum syntaktischen sowie semantisch relational-kategoriellen monoargumentalen Potential automatisch vermittelt:
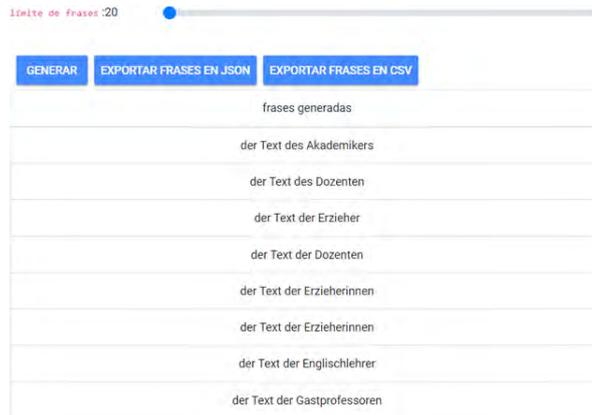
Abb. 6:   Beispiel für automatisch generierte Beispiele aus *Xera* für [DETERMINANT + TEXT + DETERMINANT IM GENITIV + ARG1[1]]

Möchte man biargumentale Schemata suchen, steht *Combinatoria* zur Verfügung (Abb. 7)[15], die sich auf die zuvor programmierten Selektionsbeschränkungen sowie syntagmatischen Relationen stützt.
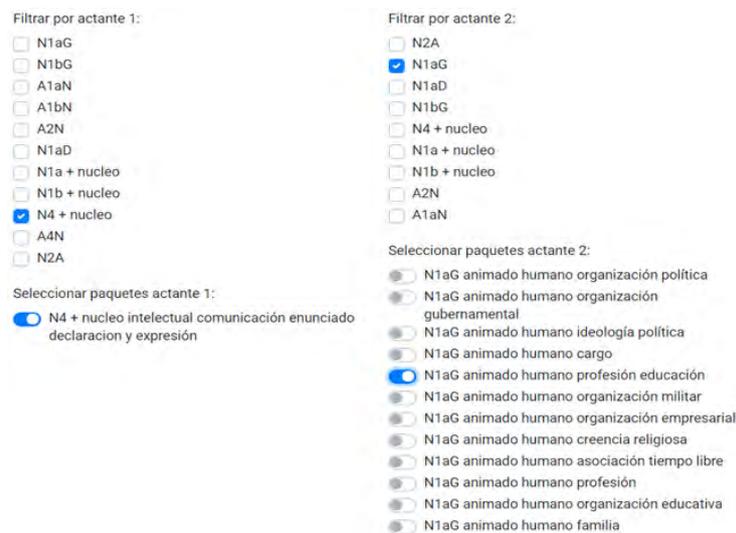
Abb. 7:   Argumentstrukturmenü bei *Combinatoria* 1.0. Beispiel TEXT

Gemäß der Auswahl (s. Abb. 7) erhält man beim Anklicken auf „Generar fra-ses" automatisch generierte Beispiele, die sich im JSON- und CSV-Format exportieren lassen:

[DETERMINANT + ARG5$_{2-+}$ TEXT + DETERMINANT IM GENITIV + ARG1$_1$]

der Bemerkungstext der Akademikerin

die Lösungstexte des Gastprofessors

der Antworttext der Englischlehrer

die Bemerkungstexte der Erzieher

der Beschreibungstext der Englischlehrerinnen

die Bemerkungstexte der Akademikerinnen

der Ankündigungstext des Gastprofessors

die Erklärungstexte des Erziehers

der Lösungstext des Dozenten

Abb. 8:   Automatisch erhobene Daten aus *Combinatoria*

4.   Zum methodologischen Verfahren angesichts eines mehrstufigen Prototypbegriffs

4.1   Einführung

Bei der Erstellung der Generatoren (s. 1.) ist einer zentralen methodologischen Frage nachzugehen, und zwar, wie man semantische Daten automatisch erhebt und verlinkt, damit sich der Lexemparadigma einer Valenzstelle automatisch generieren lässt. Erforderlich ist daher eine gründliche Auseinandersetzung damit, welche Methode bei der Auswahl der vorkommenden lexikalischen Kan-didaten oder Exemplare zwecks der Belegung der nominalen Stellen (s. Abb. 2) angewandt wird, ferner wie die semantischen Klassen (s. Abb. 5) festgelegt werden. Im Konkreten:

—   Worauf beruht die Auswahl der lexikalischen Kandidaten der jeweiligen semantischen Klassen und wie werden sie den semantischen Rollen und Argumentstellen zugeschrieben? Das heißt: Wie wird festgelegt, dass 'Dozent' eine prototypische Erscheinung bei einer ARG1$_1$ der semantischen Klasse »Beruf« beim Substantiv TEXT ist? (s. Abb. 6)

—   Worauf begründet sich die Bestimmung einer semantischen Klasse wie {ANIMADO HUMANO PROFESIÓN EDUCACIÓN} beim Substantiv TEXT?

Als Ausgangspunkt für weitere Beobachtungen gilt, dass die Festlegung semantischer Klassen sowie die Auswahl lexikalischer Kandidaten durch

mehrere ineinander greifende Beschreibungsprozeduren unterstützt wird. Hiermit stellt sich der den Generatoren zugrunde liegende Prototypbegriff als unentbehrlicher Bestandteil des methodologischen Verfahrens heraus, denn

— erstens erweist er sich bei der Auswahl sowohl der lexikalischen Kandidaten als auch der semantischen Klassen zur Leerstellenbesetzung als eine aussagekräftige Instanz heraus (s. 4.).

— zweitens gilt er als Verlinkungsfaktor mit den WordNet-Ontologien (s. 4.4.).

— drittens findet er auch bei der Bestimmung der biargumentalen Argumentstruktur, der phrasalen Umgebung und des Satzrahmens Anwendung (s. Abb. 2).

4.2   Zur Prototypdefinition

Als Prototype fassen wir typische bzw. repräsentative Instanzen für eine konkrete Argumentstruktur oder Slotbesetzung auf. Dabei handelt es sich nicht um semantische abstrakte Konzepte, da der Prototypbegriff mit der syntaktischen Argumentstruktur eines konkreten nominalen Valenzträgers sowie mit der Aktualisierung einer konkreten Bedeutungslesart zusammenhängt (kombinatorische Bedeutung von Engel 2004; vgl. auch die *classes d'objets* bei Gross 2008). Ihre Typikalität sowie Repräsentativität lässt sich auf die Interaktion valenzfundierter und frequenzbezogener Parameter stützen, wie wir noch genauer erläutern werden.

Per negationem lassen sich unsere Prototypen weiter näher bestimmen:

— Unsere Prototypen gehen nicht auf kognitionsgebundene Befunde zurück (vgl. Rosch 1978).

— Man ist nicht auf der Suche nach dem besten Exemplar einer semantischen Kategorie, d.h. man versucht nicht festzulegen, ob sich z.B. 'Kopf' als der beste Vertreter einer konkreten Klasse oder Kategorie z. B. {BELEBT MENSCHLICH KÖRPERTEIL} herausstellt. Dementsprechend wird nicht das Ziel angestrebt, die besten oder die schlechteren Vertreter der Kategorie {KÖRPERTEIL} *per se* festzulegen, indem man zentralere und periphere Kategoriebeispiele kontextlos und *per se* unterscheidet.

— Eine angemessene Beschreibung von Wortbedeutungen und darüber hinaus ihre Abgrenzung gegenüber anderen Wörtern bzw. Wortbedeutungen wird von unserem Prototypbegriff nicht geleistet, auch nicht angestrebt.

— Der Prototyp gilt nicht als eine mögliche Alternative zu Definitionen.

— Unsere lexikalischen Prototypen sind keine universellen Kategorien und ihre Festlegung geht nicht auf die Zerlegung ihrer Eigenschaften zurück.

— Die Typikalität bzw. Prototypizität wird nicht gemäß der Ähnlichkeit

eines Kandidaten mit dem "besten" Vertreter einer angegebenen Klasse festgelegt oder beschrieben.

Unterschieden werden in unserem Modell zwei zentrale Prototypenklassen: die lexikalischen Prototypen (4.3.) und die darauf aufbauenden prototypischen semantischen Klassen (4.4.).

4.3   Zum Begriff der lexikalischen Prototypen

Um es vorwegzunehmen: Lexikalische Prototypen sind konkrete prototypische Lexeme, die sich bei der Füllung einer bestimmten syntaktisch-semantischen Valenzstelle eines vorgegebenen Valenzträgers als repräsentativ bestimmen lassen. Zu ihrer Festlegung stellt sich die Wechselwirkung von zwei nicht gleichrangigen Kriterien als entscheidend heraus:

— Das ausschlaggebende Kriterium hängt mit der semantischen Belegung einer bestimmten Valenzstelle eines konkreten Valenzträgers zusammen.

— Als untergeordnetes Kriterium gilt die Häufigkeit dieses Kandidats bei Erfüllung einer semantischen Rolle in einem angegebenen phrasalen Rahmen, und dies auch unter Berücksichtigung der Argumentstrukturfrequenz (s. 4.5).

Zur Auswahl der lexikalischen Prototypen muss eingangs die Argumentstruktur des nominalen Valenzträgers beschrieben werden. Dazu lehnen wir uns an dem valenzbasierten Ansatz an, der dem multilingualen Wörterbuch *Portlex* zugrunde liegt (Domínguez Vázquez und Valcárcel Riveiro 2020). Sucht man im Falle vom Substantiv TEXT nach Kandidaten für die Slotbesetzung des Arguments `AGENS´ (`derjenige, der die Handlung durchführt´) realisiert im Muster [TEXT + DETERMINANT IM GENITIV + NOMEN], wird zunächst mittels einer CQL-Abfrage im *Korpus German Web 2013 (de TenTen13)* aus *Sketch Engine* eine Häufigkeitsliste dieser Argumentstruktur erhoben:

		Häufigkeit	Häufigkeit pro Million
1.	Text die Lied	1913	0.09658
2.	Text die Bibel	1820	0.09188
3.	Text die Buch	866	0.04372
4.	Text die Autor	844	0.04261
5.	Text die Band	837	0.04226
6.	Text die Song	755	0.03812
7.	Text die neu Testament	649	0.03276
8.	Text die alt Testament	509	0.02570

9.	Text die Petition	507	0.02560
10.	Text die Seite	487	0.02459
11.	Text die Artikel	467	0.02358
12.	Text die Mail	365	0.01843
13.	Text die Anzeige	345	0.01742
14.	Text die Rede	332	0.01676
15.	Text die heilig Schrift	331	0.01671
16.	Text die Autorin	331	0.01671
17.	Text die Urkunde	326	0.01646
18.	Text die E-Mail	324	0.01636
19.	Text die Webseite	323	0.01631
20.	Text die Evangelium	318	0.01605

Tab. 2:   20 erste Lemmatas von TEXT + Genitiv aus dem Korpus *de TenTen13*

Diese Ergebnisse werden dann gemäß semantisch-relationaler Prinzipien herausgefiltert, denn eine formal basierte Häufigkeitsliste reicht für unsere Zwecke nicht aus:

— Bei der Auswahl der lexikalischen Kandidaten besteht Bedarf an einer semantischen valenzbezogenen Analyse im Hinblick auf die relationale Bedeutung[16]. Das geht darauf zurück, dass eine gleiche Form je nach dem Valenzträger und je nach den vom ihm eröffneten Stellen, ebenfalls im Zusammenspiel mit anderen Ergänzungen, unterschiedliche semantische Rollen zum Ausdruck bringen kann. Es liegt nahe, dass die drei ersten häufigsten Beispiele in der Tab. 2 — 'Text des Liedes', 'Text der 'Bibel' und 'Text des Buches' — keine Agensrolle darstellen[17].

— Die bereits angeführten Beispiele bringen ebenfalls den Nachweis dafür, dass die Häufigkeit in unserem Beschreibungsmodell an sich kein allgemeingültiges Kriterium ist. Deutlich wird, dass „frequent" nicht immer mit „valenzbedingt" einhergeht und daher werden 'Text die Monat' (Position: 124; 88 Treffer) und 'Text die Jahr' (Position 103: 102 Treffer) aus der Kandidatenliste ausgeschlossen.

Aus dem Angeführten geht deutlich hervor, dass das Häufigkeitskriterium ausschließlich in enger Verbindung mit der Argumentstruktur, mit der semantischen Rolle und mit der aktualisierten Realisierung und Lesart eines Valenzträgers anzuwenden ist. Aus diesem Grund erstellen wir für jede monolinguale Argumentstruktur jedes Substantivs eine Liste von lexikalischen Kandidaten (grob gesagt, eine Wortschatzliste), die in paradigmatischer Beziehung zueinander stehen. Werden die lexikalischen Kandidaten herausgefiltert, erfolgt ihre kategorielle Annotation in Anlehnung an eine *ad hoc* erstellte Ontologie[18]. Somit

stellen wir fest, dass die Stelle `AGENS´ bei einem Substantiv wie TEXT in der
Regel von folgenden Substantiven besetzt wird:

Lexikalischer Prototyp	1. Stufe	2.Stufe	3. Stufe	4. Stufe
Architekt	belebt	menschlich	Beruf	
Autor	belebt	menschlich	Beruf	
Dichter	belebt	menschlich	Beruf	
Hersteller	belebt	menschlich	Beruf	
Journalistin	belebt	menschlich	Beruf	
Künstler	belebt	menschlich	Beruf	
Philosoph	belebt	menschlich	Beruf	
Sänger	belebt	menschlich	Beruf	
Schriftsteller	belebt	menschlich	Beruf	
Schüler	belebt	menschlich	Eigenschaft	
Teilnehmer	belebt	menschlich	Eigenschaft	
Verlag	belebt	menschlich	Organisation	Unternehmen
Uni	belebt	menschlich	Organisation	Bildung
EU	belebt	menschlich	Organisation	Politik
Paulus	belebt	menschlich	Eigenname	
Jury	belebt	menschlich	Kollektiv	
Papst	belebt	menschlich	Amt	

Abb. 9:   Kategorielle Beschreibung der lexikalischen Prototypen

Diese Vorgehensweise lässt somit die semantische paradigmatische Belegung
einer konkreten Valenzstelle festmachen, die zwar aus sprachwissenschaft-
licher Sicht zentral ist, jedoch angesichts der automatischen Sprachgenerierung
unerlässlich wird, denn auf die Analyse der prototypischen Kandidaten stützt
sich die Festlegung der prototypischen semantischen Klassen. Da der Program-
mierung der Generatoren nicht einzelne Wörter, sondern semantische Klassen
unterliegen, werden sie zum Schlüsselkonzept.

4.4   Zum Begriff der prototypischen semantischen Klassen

Bei der Analyse hat sich herausgestellt, dass man sich für die automatische
Datenerhebung und Sprachgenerierung (und nicht nur dafür) mit generellen
Etiketten wie [Menschen], [Tiere] oder [Gegenstand] nicht zufriedengeben

kann, da sie zu weit gefasst sind. Daher erstellen wir auf der Grundlage einer summativen Beschreibung der lexikalischen Kandidaten (Abb. 9) prototypische semantische Klassen. Ein Beispiel dafür bildet die Beschreibung der semantischen Kategorien zur Wiedergabe der Agensrolle bei TEXT (Abb. 10).
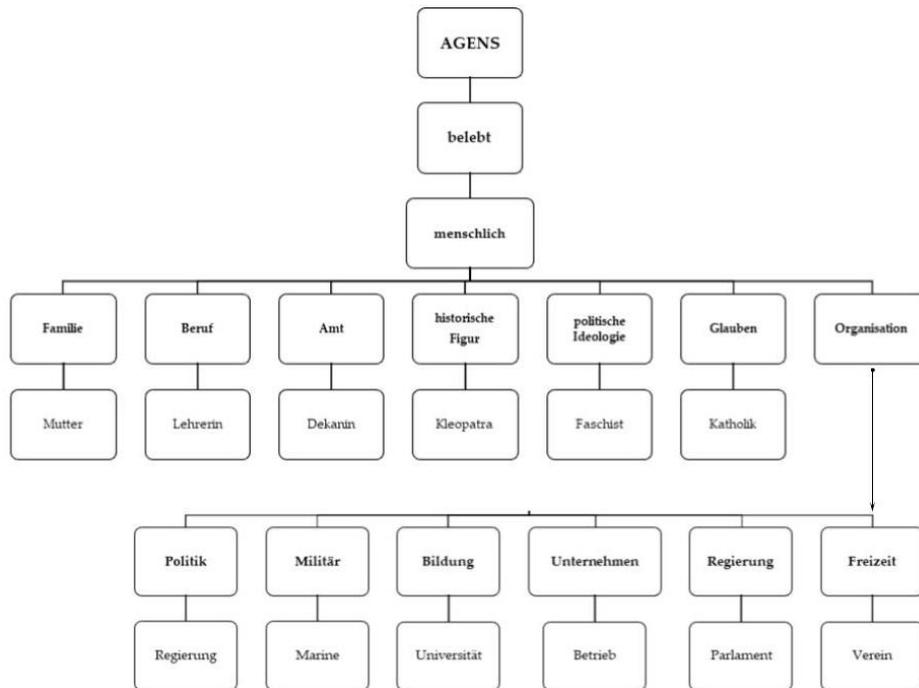
Abb. 10:   Semantische Klassen der Agensrolle bei TEXT

Aber damit ist die semantische Analyse nicht ausgeschöpft:

(a) **Zur Beziehung zwischen lexikalischen Kandidaten und semantischen Klassen:** Eine semantische Klasse mit ihren Vertretern lässt sich nicht als allgemeingültige, fixierte Kategorie für jegliche Substantive bestimmen und dementsprechend automatisch verwenden. Zur Verdeutlichung wird die semantische Klasse {BELEBT MENSCHLICH KÖRPERTEIL} herangezogen: Bei der Analyse hat sich erwiesen, dass diese semantische Klasse bei verschiedenen Substantiven, wie z.B. FARBE und SCHMERZ, als prototypisch gilt. Man kann aber dieses lexikalische *Package* — die dieser Klasse zugeordnete Information — nicht automatisch wieder verwenden, da die lexikalischen Kandidaten der semantischen Klasse {BELEBT MENSCHLICH KÖRPERTEIL} bei FARBE und SCHMERZ nicht übereinstimmen[19]:

Valenzträger	Muster	Lexikalische Prototypen	Frequenz[20]
		'Auge'	P.: 31. 460-mal
FARBE	+[DETERMINANT IM GENITIV+ADJEKTIV[FAKULTATIV]+NOMEN]	'Haar'	P.: 38. 402-mal
		'Haut'	P.: 54. 285-mal
		'Auge'	P.: 30.1949-mal
SCHMERZ	+[DETERMINANT + {KÖRPERTEIL}-+SCHMERZ]	'Haar'	P.: 263. 39-mal
		'Haut'	P.: 234. 48-mal

Tab. 3: Gegenüberstellung der Kandidaten einer semantischen Klasse bei verschiedenen Nomina. Daten aus *Sketch Engine*

Bei SCHMERZ lassen sich 'Haar' und 'Haut' im Gegenteil zu FARBE nicht als prototypische Vertreter dieser semantischen Klasse auffassen, vor allem wenn man ihre Ko-Okurrenzen den prototypischen und häufig vorkommenden Komposita gegenüberstellt (Tab. 4).

		'Kopfschmerz'	P.:1. 188948-mal
SCHMERZ	+[DETERMINANT + {KÖRPERTEIL}-+SCHMERZ]	'Rückenschmerz'	P.:2. 92363-mal
		'Bauchschmerz'	P.: 3. 45907-mal

Tab. 4: Lexikalische Prototypen bei SCHMERZ. Daten aus *Sketch Engine*

(b) **Zur Wortschatzerweiterung:** Bezüglich der Erstellung der Generatoren ist einer weiteren Frage nachzugehen, die auch an die Wichtigkeit der Festlegung semantischer prototypischer Klassen anknüpft: die Rolle der semantischen Klassen oder Kategorien, und ebenfalls der konkreten kategoriellen Merkmale ([Menschen], [Organisation]), als Verknüpfungsinstanz mit Kategorien und Sub-kategorien der WordNet-Ontologien (Gómez Guinovart und Solla Portela 2018). Dabei sind die semantischen Klassen von herausragender Bedeutung, und dies aus den folgenden Gründen:

—   Die Verlinkung auf die WordNet-Ontologien mithilfe der semantischen Klassen ermöglicht eine verfeinerte Granularität unserer Ontologie sowie eine detailliertere Beschreibung des sprachlichen Materials.

—   Dadurch gelangt man zu einer vollständigeren Wortschatzliste für jede konkrete Leerstellenbesetzung.

Dieses Verfahren der Wortschatzerweiterung trägt zur Bildung eines selegierten Lexemparadigmas mit den gleichen semantischen Eigenschaften wie denen des ursprünglichen lexikalisch-semantischen Prototyps bei. So wird die paradigmatische Achse festgelegt, die die lexikalische Auswahl bei der automatisierten Generierung von Nominalphrasen unterstützt.

Die Expandierung erfolgt dank der Anwendung der Tools *Lematiza*[21] und *Combina*[22] (Domínguez Vázquez, Solla Portela und Valcárcel Riveiro 2019). Durch ihren Einsatz können wir Wortschatzlisten aus den WordNet-Ontologien — z.B. {KÖRPERTEIL} in der Tab. 5 — gezielt zusammenstellen, aber auch semantische Klassen genauer einteilen (Tab. 6). Diese Einteilung in Klassen und Unterklassen ist enorm relevant: denn — nochmals am Beispiel vom {KÖRPERTEIL} — wir alle wissen, dass nicht alle Körperteile riechen oder wehtun, d.h. nicht alle Körperteile kommen im gleichen Prädikatausdruck vor.

WordNet: Ontologien[23]	Klassen
SUMO Ontology	BodyPart+ BodyJunction+ Organ +
Epinonyms	[1] external_body_part
Basic Level Concept	05220461-n body_part
WordNet Domains	anatomy
Top Concept Ontology	1stOrderEntity+ Living+ Part+

Tab. 5:   {KÖRPERTEIL} bei WordNet-Ontologien

Klassen	Vertreterbeispiel
[belebt] [menschlich] [Körperteil] [extern]	'Kopf'
[belebt] [menschlich] [Körperteil] [intern] [Muskel/Knochen]	'Muskel'
[belebt] [menschlich] [Körperteil] [Beschichtung]	'Haut'
[belebt] [menschlich] [Körperteil] [Organ]	'Eierstock'

Tab. 6:   Semantische Klassen mit {KÖRPERTEIL}

Bei diesem Verfahren steht der Gedanke einer gesteuerten Erhebung von einer signifikativen Anzahl an lexikalischen Kandidaten im Vordergrund. Dieser zuletzt genannte Aspekt ist nicht zu unterschätzen, denn bei der qualitativen Evaluation von automatisch generierten Daten (Hashimoto et al. 2019, Vicente et al. 2015) sind nicht nur ihre Qualität, sondern auch ihre Vielfaltigkeit zentral.

4.5   Prototypen als extern und intern vergleichende Instanzen

4.5.1 Einführung

Bei der praktischen Anwendung unseres Prototypizitätbegriffs bzw. Typikalitätsbegriffs sind wir auf unterschiedliche Fälle gestoßen, die einer näheren Beobachung bedürfen und die Prototype näher bestimmen lassen:

— Prototyp als **extern vergleichende Instanz:** Diese Auffassung ermöglicht die Gegenüberstellung eines konkreten lexikalischen Kandidats bei unterschiedlichen nominalen Valenzträgern. Ein Beispiel dafür bildet der schon angeführte Vergleich von {KÖRPERTEIL} bei FARBE und SCHMERZ (vgl. 4.4.). Eine weitere Anwendung des Prototyps als extern vergleichende Instanz kann auch bei der Gegenüberstellung verschiedener Sprachen erfolgen. Dies lässt sich beim Vorkommen der Eigennamen in den 3 beschriebenen Sprachen deutlich beobachten. Somit sind prototypisch 'Mia' oder 'Lena' im Deutschen, im Spanischen 'Susana' oder 'Patricia' und im Französischen 'Paulette' oder 'Annick'.

— Prototyp als **intern vergleichende Instanz:** Zu einem internen Vergleich erweist sich ebenfalls der Prototyp als ausschlaggebend, ferner bestätigt dieser ihn als solchen. In diesem Zusammenhang ist dem Verhältnis zwischen der Repräsentativität der Argumentstruktur, der Anzahl an lexikalischen Kandidaten sowie der Häufigkeit der lexikalischen Kookurrenzen ein zentraler Stellenwert zuzuweisen. Darauf wird in den nachstehenden Abschnitten eingegangen.

4.5.2 Prototypen: Argumentstruktur und Kandidantenliste

Als intern vergleichende Instanzen lassen sich die Prototypen nach verschiedenartigen Prototypizitätstypen oder -grade auf einer fiktiven Prototypizitätskala unterscheiden:

(i) Als prototypischste Argumentstruktur stellt sich ein repräsentatives und häufig vorkommendes Muster mit einer breiten Palette an verschiedenen lexikalischen Kandidaten heraus, die außerdem häufig vorkommen. Ein Beispiel dafür bildet die Argumentstruktur [DISKUSSION + Präposition[über/um] + Nomen] bei Wiedergabe der semantischen Rolle `THEMA´:

	Häufigkeit	*Scores*
Diskussion über	**114,929**	**6.57%**
Diskussion um	110,558	6.32%
Diskussion zu	47,816	2.73%
"Diskussion" außerhalb + noun	21	0,0%
"Diskussion" anlässlich + noun	15	0,0%
Diskussion bzgl.	13	0,0%

Tab. 7: Gegenüberstellung von Frequenzangaben bei DISKUSSION + Präposition

Die Tabelle gibt zu erkennen, dass das Muster [DISKUSSION + Präposition[über] + Nomen] häufiger als andere Kombinationen vorkommt und dass

es im Vergleich zu anderen möglichen Realisierungsmustern, z.B. [Komposita + -DISKUSSION], bei Wiedergabe derselben semantischen Rolle repräsentativ ist. Hinzu kommt, dass sie über unterschiedliche und häufig vorkommende lexikalische Kandidanten verfügt:

Diskussion über	Frequenz
Thema	5,189
Zukunft	2,886
Sinn	1,893
Frage	2,383
Rolle	874
Problem	874
Einführung	714
Umgang	743
Inhalt	765
Möglichkeit	838

Tab. 8:   Lexikalische Kandidaten nach Häufigkeitsangaben

(ii) Im Gegensatz zu (i) sind zwar nicht häufige Argumentmuster vorhanden, die sich aber als repräsentativ herausstellen: Aus einer valenzbasierten Analyse der aus *Sketch Engine* erhobenen Daten über das Muster [Adjektiv + SCHMERZ] lässt sich ableiten[24], dass nicht valenzbezogene Realisierungen bezüglich der Gesamtanzahl der Adjektive als auch der Ko-Okurrenzen in der Mehrheit stehen. Somit zeigt eine semantisch basierte Analyse der ersten 100 Substantive, dass 92% der vorkommenden Adjektive nicht valenzgefordert sind, bei den restlichen 8% handelt es sich um Ergänzungen.
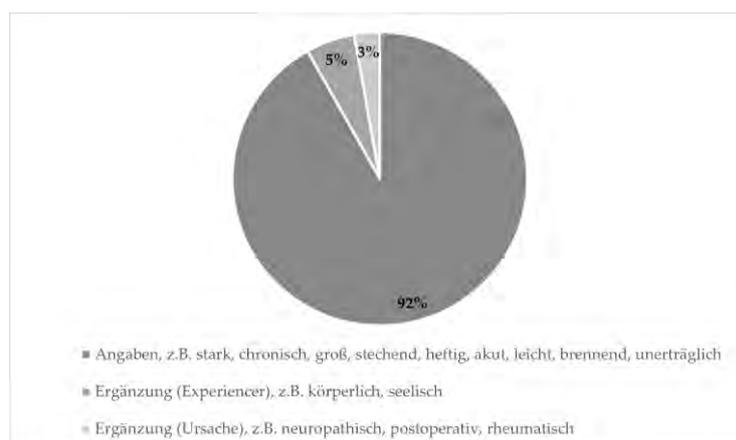
Graphik 1:   Semantisch valenzbasierte Analyse von Adjektiven bei SCHMERZ

In Bezug auf die Häufigkeit der valenzbezogenen Realisierungen lässt sich das Muster zweifelsohne nicht als häufig bezeichnen. Indem es sich aber als repräsentativ für eine mögliche Realisierung der semantischen Rolle `EXPERIENCER´ erweist, wird es in den Generatoren berücksichtigt.

Hiermit lässt sich ebenfalls beobachten, dass dieses Muster eine sehr eingeschränkte Kandidatenliste zeigt, deren Vertreter — 'körperlich' (5,899 Treffer; 5. Häufigkeitsposition) und 'seelisch' (10. Häufigkeitsposition mit 3,621 Treffer) — aber ganz oben in der Frequenzliste stehen.

		Häufigkeit	Häufigkeit pro Million
1	stark Schmerz	30,125	1.52
2	chronisch Schmerz	21,934	1.11
3	groß Schmerz	12,875	0.65
4	stechend Schmerz	8,063	0.41
5	körperlich Schmerz	5,899	0.30
6	heftig Schmerz	5,165	0.26
7	akut Schmerz	5,045	0.25
8	leicht Schmerz	4,624	0.23
9	brennend Schmerz	3,673	0.19
10	seelisch Schmerz	3,612	0.18

Tab. 9:   Häufigkeitsangaben von Adjektiven + SCHMERZ

(iii) Nicht repräsentative wie auch nicht häufig vorkommende Argumentstrukturen mit wenigen und nicht frequenten Kandidaten werden nicht in die Generatoren integriert. Zur Veranschaulichung derartiger Fälle dient das Muster [SCHMERZ + Genitiv: {BELEBT MENSCHLICH KÖRPERTEIL}]. Hingegen deuten die Korporadaten zur Prototypizität der Kompositarealisierung hin:

	Häufigkeit	Häufigkeit pro Million
107 Schmerz die Kopf	21	< 0.01
1 Kopfschmerz	188950	954

Tab. 10:   Gegenüberstellung von 'Kopf' in Verbindung mit SCHMERZ

4.5.3 Prototypen: semantische Klassen und Argumentstruktur

Hinsichtlich des Verhältnisses zwischen semantischen Klassen und Argumentmustern sind zwei Fälle zu beachten:

(i)   Integriert werden in die Generatoren semantische Klassen mit einer signi-fikativen Anzahl an Vertretern, die sich einer häufigen und repräsentati-ven Argumentstrukturmuster zuschreiben lassen. Als Beispiel stellt sich die semantische Klasse {BELEBT MENSCHLICH KÖRPERTEIL} in Form eines Kompositumsbestandteils bei SCHMERZ heraus:
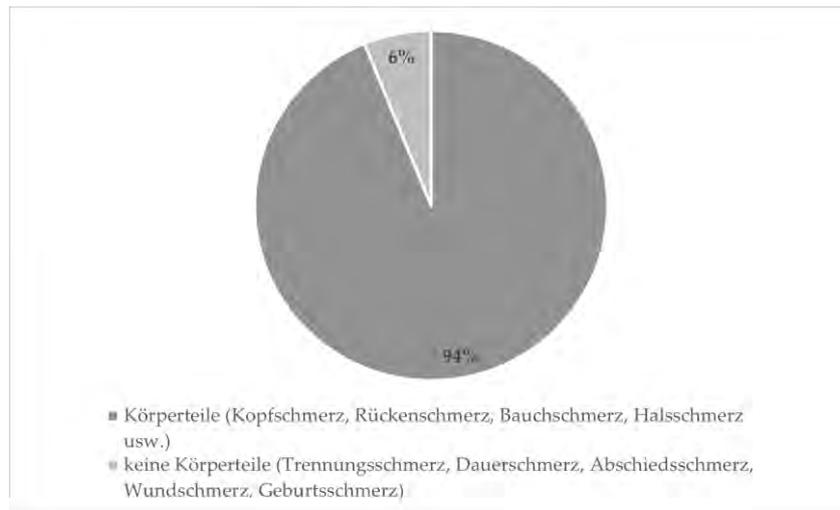
Graphik 2:   Prozentansätze von Komposita mit Körperteil bei SCHMERZ

(ii) Im Gegensatz zu (i) stößt man auch auf Fälle von semantischen Klassen mit wenigen aber häufig vorkommenden Vertretern. Ein Beispiel dafür findet das syntaktische Muster [FARBE + Genitiv], dessen Genitivrealisie-rung mit vielen semantischen Klassen ausgestattet ist. Darunter lässt sich als Beispiel dafür die semantische Klasse {MATERIELL GEGENSTAND SCHÖNHEITSPLEGE KOSMETIK} nennen, die eine eingeschränkte Anzahl an Vertretern hat — z.B. 'Lippenstift' — , die aber schon sehr häufig auf-tauchen.

4.5.4   Prototypen in den biargumentalen Strukturen

Bisher war von monoargumentalen Strukturen die Rede. Deutlich werden sollte allerdings unmittelbar, dass sich bei der Entfaltung des im Nomen enhal-tenen Potentials nicht alle Kombinationen und Selektionsbeschränkungen vor-aussehen bzw. voraussetzen lassen. Im Hinblick darauf liegt es nahe, dass nicht alle Vertreter jeder semantischen Klasse eines Arguments[X] mit allen Vetretern einer anderen Klasse eines Arguments[Y] in einer biargumentalen Realisierung kombinierbar sind.

Bei der Festlegung biargumentaler Muster greift man wieder auf den Prototypbegriff zurück. Somit ist es erforderlich, zuerst das Vorhandensein des biargumentalen Musters im Korpus zu prüfen. Liegt der Analyse das Muster [Determinant + {Adjektiv} + ANTWORT + Determinant im Genitiv + Argument N1 + auf + Determinant im Akkusativ + Argument N2] zugrunde, lässt sich festhalten, dass die semantische Klasse {ORGANISATION REGIERUNGSGEBUN-DEN} für N1 und die semantische Klasse {INTELLEKTUELLES KOMMUNIKA-TION} für N2 prototypisch sind. Beachtet man konkret die vorkommenden lexikalischen Prototypen, ist zu beobachten, dass die lexikalischen Kandidaten 'Anfrage' und 'Frage' in der Mehrheit stehen:

	die Bundesregierung auf eine Anfrage	465
	die Landesregierung auf eine Anfrage	92
	die Senat auf eine Anfrage	81
	die Bundesregierung auf die Anfrage	66
	die Landesregierung auf die Anfrage	38
ANTWORT	die Regierung auf eine Anfrage	37
	die Verwaltung auf eine Anfrage	35
	die Bundesregierung auf die Frage	35
	die Stadtverwaltung auf eine Anfrage	31
	die Verwaltung auf die Anfrage	29

Tab. 11:   Biargumentales Muster von ANTWORT

Zum Erlangen akzeptabler Ergebnisse bei der sprachlichen Generierung ist zwar eine vertiefte Analyse der biargumentalen Muster unentbehrlich, jedoch kann jede mögliche semantisch-syntaktische Kombination nicht manuell analysiert werden. Aus diesem Grund lassen sich die Resultate unseres kombinatorischen Tools — *Combinatoria* — mithilfe der prädiktiven Methode *Word2vec* (Mikolov et al. 2013) auf die Kompatibilität der lexikalischen Kandidaten jedes Arguments überprüfen (Domínguez Vázquez 2020). Angestrebt wird dabei nicht, die Bedeutung eines Wortes einem anderen gegenüberzustellen — wie beim *Embedding Viewer* von *Sketch Engine* oder die möglich abrufbar paradigmatische Datenerhebung aus *Derekovecs*[25], sondern die Beseitigung von Inkompatibilitäten auf der Ebene der kombinatorischen Bedeutung. Dieses Vorgehen trägt zur semantischen Kohärenz bei, einem Qualitätsfaktor bei jedem Generator natürlicher Sprachen. Der Zugriff auf *word embeddings* wird auch die Integration von nicht valenzbezogenen Adjektiven in die nominalen Argumentmuster ermöglichen, damit der phrasale Kontext vermittelt wird: *Der {unangenehme} | {angenehme} | {üble} | {intensive} Geruch.*

5.   Auf dem Weg zum automatischen Wörterbuch

Bei den vorangehenden Abschnitten ist auf die Integration und Rückkoppe-
lung verschiedener Arbeitsschritte und Beschreibungsebenen zwecks der auto-
matisierten Datenerhebung und -generierung von syntaktisch-semantischen
Valenzmustern eingegangen. Der Fokus hat insbesondere auf der Entwicklung
eines mehrsichtigen Prototypbegriffs gelegen, denn dieser stellt sich für die
Verlinkung zwischen zwei zentralen Beschreibungsebenen als ausschlaggebend
heraus: zum einem die automatische Erhebung semantischer Daten, zum anderen
die Generierung des syntaktisch-semantischen Entfaltungspotentials der Nomi-
nalphrase.

Die Sprachgeneratoren, *Xera* und *Combinatoria*, können als selbständige
Tools, als lexikalische bzw. sprachliche Informationssysteme (s. Villa Vigoni-
Thesen 2018) gebraucht werden. Darüber hinaus fassen wir sie als innovative
Modelle für automatische und interaktivere Valenzwörterbücher bzw. syntag-
matische Kombinationswörterbücher auf. Sie bringen gegenüber von aktuellen
online syntagmatischen Wörterbüchern gewisse Neuerungen mit sich:

Die semantischen Rollen, die prototypischen semantischen Klassen und
die lexikalischen Prototypen gelten bei der Erstellung und bei der Hanhabung
der Ressource als entscheidende Bausteine. Insofern bieten die Generatoren
nicht nur eine ontologisch fundierte Beschreibung, sondern auch ein Inventar
an semantischen Kategorien und an lexikalischen Kandidaten zur Belegung
einer konkreten Valenzstelle an. Somit leisten sie, wie klassische Valenzwörter-
bücher, Hilfe in einer typischen Produktionssituation.

Aus typologischer Sicht bestimmen zwei Schlüsselkonzepte, automatisch
und interaktiv, den aktuellen Stand der Prototypen:

— Es handelt sich um automatische Ressourcen in dem Sinne, dass sowohl
   für die Datenerhebung als auch für die Datengenerierung Techniken und
   Verfahren aus der NLP und NLG eingesetzt bzw. neu konzipiert wurden.
   Die Generatoren können eine Grundlage für die Entwicklung künftiger
   automatischer bzw. automatisch computergestützer Valenzwörterbücher
   bilden.

— Beide Ressourcen sind ebenfalls als interaktive Tools konzipiert, indem im
   Gegenteil zu früheren syntagmatischen Werken eine Interaktionshand-
   lung zwischen Werk und Nutzer sttatfindet. Bei ihrer Handhabung kann
   in ihrem aktuellen Stand eine informativ-beschreibende Interaktion erfolgen.
   Folglich kann man zur Vergewisserung in einer Produktions- oder Korrek-
   tursituation gezielt bestimmte Muster bzw. Kombinationen abfragen.
   Zwei interaktive Herangehensweisen befinden sich in einer Entwicklungs-
   phase: (i) eine experimentativ-nachschlagende Interaktion (Selbstgenerie-
   rung und -überprüfung eines Ausdrucks) und (ii) eine experimentativ–
   betreute Interaktion (Durchführung der von den Tools angebotenen online
   Übungen). Insgesamt sind die didaktischen Anwendungen der Generato-

ren vielfältig (Domínguez Vázquez, Sanmarco Bande, Solla Portela und Valcárcel Riveiro 2019: 133-135).

Neben der Anwendung der Generatoren als selbständige Ressourcen und als Modelle für neuartige Valenzwörterbücher ist ein weiterer Anwendungsbereich zu ergänzen: die Integration ihrer Daten in andere E-Tools sowie die Anwendung der der Entwicklung der Generatoren zugrundeliegenden Tools für andere Informationssysteme. Beide sind wichtige Aufgaben der elektronischen Lexikographie:

> Another aspect is smart use and reuse of dictionary content. Namely, dictionaries often remain isolated entities, whereas the user needs and habits indicate that it would be much more useful to have them linked to other dictionaries and language resources, or even integrated in various tools. (siehe Elex 2019[26])

6.   Ausblick

Eine ausführliche Analyse der syntaktisch-semantischen Schnittstelle ist aus linguistischer Perspektive zentral, da die konkrete Besetzung der vom Valenzträger eröffneten Stellen zur Prädikatsinterpretation und darüber hinaus zur Abgrenzung gegenüber anderen Prädikaten führt. Aus Programmierungssicht ist ebenfalls die semantische Information unentbehrlich, denn Maschinen verfügen nicht über ein semantisches Wissen, daher findet in unserem Modell eine syntaktisch-semantisch valenzfundierte Beschreibung des vom Valenzträger festgelegten syntaktisch-semantischen Rahmens statt. Im Hinblick auf die notwendige syntagmatische und paradigmatische Selektion der von Generatoren gelieferten Daten, ist ein mehrstufiger Prototypbegriff entwickelt worden, der sich als wesentlicher Bestandteil des methologischen Verfahrens erweist. Sowohl aus der Entwicklungsperspektive als auch aus der Nutzersicht rückt in unserem Ansatz die Darlegung der kombinatorischen Bedeutung auf phrasaler Ebene in den Vordergrund. Die automatische Generierung von satzwertigen Realisierungen stellt sich als nächstes Ziel heraus.

Beide schon einsatzbereite Prototype verkörpern einen typologischen Vorschlag zu neuen automatischen computergestützten syntagmatischen pluringualen Wörterbüchern, die ebenfalls unterschiedliche Interaktionen beachtet.

Insgesamt zeichnet der Beitrag nach, wie neue lexikographische Tools auf der Grundlage des schon vorhandenen lexikograpischen Wissens, der Interaktion zwischen der Lexikographie und der natürlichen Sprachverarbeitung (NLP) und -generierung (NLG) sowie der Datenintegration und Ressourceninteroperabilität zustande kommen können.

Danksagung

Die Ergebnisse dieser Forschung stehen im Zusammenhang mit dem Untersuchungsvorhaben „Mehrsprachige Generierung von nominalen Argumentstrukturen und automatische syntaktisch-semantische Datenerhebung", gefördert vom Programm „Ayudas Fundación BBVA a Equipos de Investigación Científica 2017", sowie mit dem Forschungsprojekt „Mehrsprachige Generierung von nominalen Argumentstrukturen mit Anwendung bei der Produktion in Fremdsprachen", gefördert von FEDER/spanischem Ministerium für Wirtschaft, Industrie und Wettbewerb — Staatliche Forschungsagentur (FFI2017–82454-P; Programm Exzellenz).

Endnoten

1. Die verwendeten Korpora bei der Erstellung dieses Wörterbuchs sind: für Deutsch *COSMAS II* (https://www.ids-mannheim.de/cosmas2/web-app/), für Französisch *FRANTEXT* (https://www.frantext.fr/), für Galicisch CORGA (http://corpus.cirp.gal/corga), für Italienisch *PAISÀ* (https://www.corpusitaliano.it/) und für Spanisch *CREA* (http://corpus.rae.es/ creanet.html). In einigen Fällen haben wir auf *Sketch Engine* oder das Web als Korpus zugegriffen.

2. Ressourcen, die syntaktisch-semantische Informationen anbieten, stehen zur Verfügung. Ein Beispiel dafür bilden *Propbank*, *CPA* (Hanks und Pustejovsky 2005) oder *AnCora*.

3. Sie sind unter dem Link http://portlex.usc.gal/combinatoria abrufbar.

4. Siehe http://adimen.si.ehu.es/web/MCR.

5. Andere Autoren wie Kilgarriff et al. (2008) haben sich mit der Findung guter Wörterbuchbeispiele im Korpus auseinandergesetzt. Hervorzuheben sei hier, dass unsere Belege das valenzfundierte Entfaltungspotential und seine Varietät aufzeigen müssen, nur angesichts der Ressourcetypologie lassen sie sich dann als gute Beispiele bezeichnen.

6. Alle Frequenz- und Korpusdaten, auf die man sich im Beitrag bezieht, sind den Korpora aus *Sketch Engine* entnommen. Im Konkreten sind die Daten für das Deutsche dem *Korpus German Web 2013 (de TenTen13)* entnommen.

7. Im Konkreten: *Sumo Ontologie* (Niles und Pease 2001), *Epinonyms* (Gómez Guinovart und Solla Portela 2018), *Basic Level Concept* (Izquierdo et al. 2007), *WordNet Domains* (Bentivogli et al. 2004) und *Top Concept Ontology* (Álvez et al. 2008).

8. Bei Arbeiten wie Renau und Nazar (2016) geht es um die automatische Mustererhebung (*pattern*), nicht um die Generierung.

9. In der Graphik wird auf unterschiedliche Ressourcen und *Tools* Bezug genommen, die bei den verschiedenen Entwicklungsphasen der Generatoren verwendet werden. Voneinander abzugrenzen sind bereits vorhandene Tools — wie das Korpustool *Sketch Engine*, das *Freeling Tagger* (Padró et al. 2010) oder semantische Netze, wie *WordNet* (Gómez Guinovart und Solla Portela 2018*)* — gegenüber denen, die für das Forschungsvorhaben *ad hoc* entwickelt wurden, — wie *APIs* und die Tools *Lematiza*, *Combina*, *Flexiona* (siehe dazu Domínguez Vázquez, Solla Portela und Valcárcel Riveiro 2019). Frei zugänglich sind alle unsere Ressourcen bis auf eine Ausnahme: der Generator des Satzrahmens, der gerade getestet wird.

10. Für weitere Beschreibungsphasen siehe 4.

11. Da die Generatoren ebenfalls Daten fürs Spanische und Französische anbieten, sind ebenfalls die Argumentstellen in diesen Sprachen einzeln beschrieben. Für weitere Details siehe

Domínguez Vázquez und Valcárcel Riveiro (2020) und Valcárcel (2017: 193). Auflösung der Abkürzungen: GS steht für Genitivus Subiectivus, PrpE für Präpositivergänzung und NmlE für Nominalergänzung.

12. Die in der Tabelle angeführten kategoriellen Beschreibungskategorien werden bei der Entwicklung der Prototypen weiter spezifischer differenziert, hier gelten sie als Ausgangspunkt. Vgl. dazu auch Martín Gascueña und Sanmarco Bande (2019).

13. Es besteht die Möglichkeit, die Daten einer bestimmten semantischen Klasse, unterschiedlicher Klassen oder aller Klassen gleichzeitig abzurufen.

14. Übersetzung im Deutschen: {BELEBT MENSCHLICH BERUF AUSBILDUNG}.

15. Da die Selektion der Argumente im Auswahlmenü für Laien Schwierigkeiten bereiten kann, werden zur Zeit andere Zugriffstrukturen getestet. Ein Beispiel dafür bietet den Vorschlag zu einem ontologisch-semantischen Datenzugriff im Tool *XeraWord* (Domínguez Vázquez et al. 2020), das auf der Grundlage von *Xera* (Domínguez Vázquez et al. 2020) für das Galicische und Portugiesische entwickelt worden ist.

16. Eine semantisch verankerte Datenerhebung ist bei den verwendeten Korpora — wie bei vielen anderen — nicht möglich, da sie nicht semantisch annotiert sind.

17. Es handelt sich eigentlich um das Argument — `dasjenige, das etwas hat/über etwas verfügt´ —, zwar eine Ergänzung, aber kein `AGENS´ (vgl. Abb. 2).

18. Die erstellte Ontologie vereinbart vorhandene Inventare der Valenzlexikographie, Daten aus WordNet sowie ein *bottom-up* -Verfahren.

19. Da die einer semantischen Kategorie zugeschriebene Kandidatenliste bei allen beschriebenen Substantiven nicht gleich sein muss bzw. nicht ist, müssen die erstellten Kandidantenlisten bei jedem Substantiv geprüft werden.

20. P. steht für Position.

21. Die Lexeme, die wir aus den Korpora in *Sketch Engine* erhoben haben, bietet *Lematiza* begleitet von ihren jeweiligen *synsets* und verlinkt auf die verschiedenen WordNet-Ontologien automatisch an.

22. *Combina* ermöglicht kombinierte Abfragen an die WordNet-Ontologien sowie die gezielte Erhebung lexikalischer Kandidaten mit ihren *synsets*.

23. Zu den WordNet-Ontologien siehe Fußnote 7.

24. Es herrscht kein Konsens darüber, ob Adjektive sowie Kompositabestandteile als Ergänzungen fungieren können. Hier wird der Ansatz vertreten, dass die syntaktisch-semantische Funktion nicht von der Form bestimmt wird und dementsprechend diese Formen auch die Funktion einer Ergänzung erfüllen können.

25. Unter: http://corpora.ids-mannheim.de/openlab/derekovecs/.

26. Unter: https://elex.link/elex2019/.

Literaturverzeichnis

Álvez, J., J. Atserias, J. Carrera, S. Climent, E. Laparra, A. Oliver und G. Rigau. 2008. Complete and Consistent Annotation of WordNet Using the Top Concept Ontology. Calzolari, N. et al. (Hrsg.). 2008. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), 28–30 May 2008, Marrakech, Morocco*: 1529-1534. Paris: European Language Resources Association (ELRA).

Bardanca Outeiriño, D. 2020. *Automatic Generation of Dictionaries*. Master Thesis: Universidade de Santiago de Compostela.

Bentivogli, L., P. Forner, B. Magnini und E. Pianta. 2004. Revising the WordNet Domains Hierarchy: Semantics, Coverage and Balancing. Sérasset , G. et al. 2004. *Proceedings of the COLING 2004 Post Conference Workshop on Multilingual Linguistic Resources (MLR2004), August 28th, 2004, University of Geneva, Switzerland*: 101-108. Geneva: COLING.

Domínguez Vázquez, M.J. 2020. Aplicación de WordNet e de word embeddings no desenvolvemento de prototipos para a xeración automática da lingua. *Linguamática* 12(2): 71-80.

Domínguez Vázquez, M.J, M.T. Sanmarco Bande, M.A. Solla Portela und C. Valcárcel Riveiro. 2019. Die Einsetzung von neuartigen online Werkzeugen im Fremdsprachenunterricht: multilinguale automatische Generatoren des nominalen Kombinationspotentials. *Aussiger Beiträge* 13(2019). *Germanistische Shriftenreihe aus Forschung und Lehre. Acta universitatis purkynianae. Facultatis philosophicae studia germanica:* 119-158.

Domínguez Vázquez, M.J., M.A. Solla Portela und C. Valcárcel Riveiro. 2019. Resource Interoperability: Exploiting Lexicographic Data to Automatically Generate Dictionary Examples. Kosem, I. et al. (Hrsg.). 2019. *Electronic Lexicography in the 21st Century: Smart Lexicography. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 51-71. Brno: Lexical Computing.

Domínguez Vázquez, M.J. und C. Valcárcel Riveiro. 2020. PORTLEX as a Multilingual and Crosslingual Online Dictionary. Domínguez Vázquez M.J., M. Mirazo Balsa und C. Valcárcel Riveiro (Hrsg.). 2020. *Studies on Multilingual Lexicography*: 135-158. Berlin/Boston: De Gruyter.

Engel, U. 1996. Semantische Relatoren. Ein Entwurf für künftige Valenzwörterbücher. Weber, N. (Hrsg.). 1996. *Semantik, Lexikographie und Computeranwendungen*: 223-236. Tübingen: Niemeyer.

Engel, U. 2004. *Deutsche Grammatik — Neubearbeitung*. München: Iudicium.

Geyken, A., F. Wiegand und K.-M. Würzner. 2017. On-the-fly Generation of Dictionary Articles for the DWDS Website. Kosem, I. et al. (Hrsg.). 2017. *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of the eLex 2017 Conference, 19–21 September 2017, Leiden, the Netherlands*: 560-570. Brno: Lexical Computing.

Gao, J. und H. Liu. 2020. Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners. *Lexikos* 30: 111-142.

Gómez Guinovart, X. und M.A. Solla Portela. 2018. Building the Galician Wordnet: Methods and Applications. *Language Resources and Evaluation* 52(1): 317-339.

Gross, G. 2008. *Les classes d'objets*. Paris: Presses de l'Ecole normale supérieure.

Hanks, P. und J. Pustejovsky. 2005. A Pattern Dictionary for Natural Language Processing. *Revue Francaise de Linguistique Appliquée* 10(2): 63-82.

Hashimoto, T.B, H. Zhang und P. Liang. 2019. Unifying Human and Statistical Evaluation for Natural Language Generation. Burstein, J. et al. (Hrsg.). 2019. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*: 1689-1701. Minneapolis: Association for Computational Linguistics.

Héja, E. und D. Takács. 2012. Automatically Generated Online Dictionaries. Calzolari, N. et al. (Hrsg.). 2012. *Proceedings of the Eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey, May 21–27, 2012 (LREC 2012):* 2487-2493. Istanbul: European Language Resources Association.

Izquierdo, R., A. Suárez und G. Rigau. 2007. Exploring the Automatic Selection of Basic Level Concepts. Angelova, G. et al. (Hrsg.). 2007. *Proceedings of the International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, September 27–29, 2007 (RANLP'07)*: 298-302. Borovets: Incoma Ltd.

Kabashi, B. 2018. A Lexicon of Albanian for Natural Language Processing. Čibej, J. et al. (Hrsg.). *2018. Proceedings of the 18th EURALEX International Congress: Lexicography in Global Contexts, Ljubljana, July 17–21, 2018:* 855-862. Ljubljana: Ljubljana University Press, Faculty of Arts.

Kilgarriff, A., M. Husák, K. McAdam, M. Rundell und P. Rychlý. 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, E. und J. DeCesaris (Hrsg.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 425-432. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Kosem, I., I. Koppel, T. Zingano Kuhn, J. Michelfeit und C. Tiberius. 2019. Identification and Automatic Extraction of Good Dictionary Examples: The Case(s) of GDEX. *International Journal of Lexicography* 32(2): 119-137.

Laufer, B. und I.S.P. Nation. 2012. Vocabulary. Gass, S. und A. Mackey (Hrsg.). 2012. *The Routledge Handbook of Second Language Acquisition*: 163–176. London/New York: Routledge.

Martín Gascueña, R. und M.T. Sanmarco Bande. 2019. Tratamiento lexicográfico de las relaciones de meronimia. Calvo Rigual, C. und F. Robles Sabater (Hrsg.). 2019. *La investigación en lexicografía hoy. Diccionarios bilingües, Lingüística y uso del diccionario. Vol. I*: 307-323. Valencia: Universitat de València, Servei de Publicacions.

Mikolov, T. et al. 2013. Efficient Estimation of Word Representations in Vector Space. Bengio, Y. und Y. LeCun (Hrsg.). 2013. *Proceedings of the First International Conference on Learning Representations (ICLR 2013), Scottsdale, Arizona, USA, May 2–4, 2013, Workshop Track Proceedings*: 1301-3781. Scottsdale, Arizona.

Nallapati, R. et al. 2016. Abstractive Text Summarization Using Sequence-to-sequence RNNs and Beyond. Riezler, S. und Y Goldberg (Hrsg.). 2016. *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL), Berlin, Germany, August, 11–12, 2016*: 280-290. Berlin: Association for Computational Linguistics.

Nied Curcio, M. 2014. Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht. Abel, A. et al. (Hrsg.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, Bolzano/Bozen, Italy, 15–19 July 2014*: 263-280. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.

Niles, I. und A. Pease. 2001. Towards a Standard Upper Ontology. Zúñiga, G. (Hrsg.). 2001. *FOIS '01. Proceedings of the International Conference on Formal Ontology in Information Systems*: 2-9. New York: ACM.

Otter, D.W., J.R. Medina und J. Kalita. 2020. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*: 1-21. Doi: 10.1109/TNNLS.2020.2979670.

Padró, L., M. Collado, S. Reese, M. Lloberes und I. Castellón. 2010. FreeLing 2.1: Five Years of Open-source Language Processing Tools. Calzolari, N. et al. (Hrsg.). 2010. *Proceedings of the Seventh International Conference on Language Resources and Evaluation, Valetta, Malta, May 17–23, 2010 (LREC 2010)*: 931–936. La Valleta: European Language Resources Association.

Renau, I. und R. Nazar. 2016. Automatic Extraction of Lexical Patterns from Corpora. Margalitadze, T. und G. Meladze (Hrsg.). 2016. *Proceedings of the XVII EURALEX International Con-*

gress: Lexicography and Linguistic Diversity, Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia, 6–10 September, 2016: 823-830. Tbilisi: Ivane Javakhishvili Tbilisi State University.

Roemmele, M. 2016. Writing Stories with Help from Recurrent Neural Networks. Schuurmans, D. und M. Wellman (Hrsg.). 2016. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*: 4311-4312. Phoenix: AAAI Press.

Rosch, E. 1978. Prototype Classification and Logical Classification: The Two Systems. Scholnik, E.K. (Hrsg.). 1978. *New Trends in Cognitive Representation: Challenges to Piaget's Theory*: 73-86. Hillsdale, NJ: Erlbaum.

Schumacher, H. 2006. Deutschsprachige Valenzwörterbücher. Vilmos A. et al. (Hrsg.). 2006. *Dependenz und Valenz. Ein internationales Handbuch der zeitgenössischen Forschung*: 1396-1424. Berlin/ New York: De Gruyter.

Sordoni, A. et al. 2015. A Neural Network Approach to Context-sensitive Generation of Conversational Responses. Mihalcea, R. et al. (Hrsg.). 2015. *Proceedings of the 2015 Annual Conference of the North American Chapter of the Association for Computational Technologies (ACL): Human Language Technologies*: 196-205. Denver: Association for Computational Linguistics.

Valcárcel, C. 2017. Las construcciones N1N2 como realizaciones actanciales del sustantivo en francés y su tratamiento en el diccionario multilingüe PORTLEX. Domínguez Vázquez, M.J. und S. Kutscher. (Hrsg.). 2017. *Interacción entre gramática, didáctica y lexicografía: estudios contrastivos y multicontrastivos*: 193-207. Berlin: De Gruyter.

Vicente, M. et al. 2015. La generación de lenguaje natural: análisis del estado actual. *Computación y Sistemas* 19(4): 721-756.

Villa Vigoni-Thesen. 2018. *Wörterbücher für die Zukunft — Die Zukunft der Wörterbücher*. https://www.emlex.phil.fau.de/andere-publikationen/.

Online-Ressourcen

AnCora. URL: http://clic.ub.edu/corpus/es/ancora.

Combinatoria = María José Domínguez Vázquez (dir.), Carlos Valcárcel Riveiro, Daniel Bardanca Outeiriño, Natalia Catalá Torres, Nerea López Iglesias, Rosa Martín Gascueña, Mónica Mirazo Balsa, María Teresa Sanmarco Bande und Laura Pino Serrano. 2020. *Combinatoria. Prototipo online para la generación biargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela.
URL: http://portlex.usc.gal/combinatoria/

Portlex = María José Domínguez Vázquez (dir.), Carlos Valcárcel Riveiro, Mónica Mirazo Balsa, María Teresa Sanmarco Bande, Alberto Simões und Mário João Vale. 2018. *Portlex. Diccionario multilingüe de la valencia del nombre*. Universidade de Santiago de Compostela.
URL: http://portlex.usc.gal/portlex/

Propbank. URL: http://verbs.colorado.edu/~mpalmer/projects/ace.html

Sketch Engine. URL: https://www.sketchengine.eu

Xera = María José Domínguez Vázquez (dir.), Carlos Valcárcel Riveiro, Daniel Bardanca Outeiriño, José Antonio Calañas Continente, Natalia Catalá Torres, Rosa Martín Gascueña, Mónica Mirazo Balsa, María Teresa Sanmarco Bande und Laura Pino Serrano. 2020. *Xera. Prototipo online para la generación automática monoargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela.

URL: http://portlex.usc.gal/combinatoria/usuario

XeraWord = María José Domínguez Vázquez (dir.), Daniel Bardanca Outeiriño, María Caíña Hurtado, Xavier Gómez Guinovart, José Luis Iglesias Allones, Alberto Simões, Carlos Valcárcel Riveiro, María Álvarez de la Granja und Francisco A. Cidrás Escaneo. 2020. *XeraWord. Prototipo de xeración automática da argumentación da frase nominal en galego e portugués*. Santiago de Compostela: Instituto da Lingua Galega.
URL: http://ilg.usc.gal/xeraword/

Literary Quotations in Bilingual Dictionaries: A Case Study of a Nineteenth-century Dutch–Chinese Dictionary

Audrey Heijns, *School of Foreign Languages, Shenzhen University, Guangdong, China (ajheijns@gmail.com)*

Abstract: The phenomenon of literary quotations in dictionaries finds its origin in the days when the search for equivalents was intertextual rather than interlingual based. For both monolingual and bilingual dictionaries the purpose of literary quotations is to supply evidence of usage and illuminate the meaning of the relevant headword, with the difference that for bilingual dictionaries this is done via translation. This article examines this phenomenon in a nineteenth century Dutch–Chinese dictionary by analysing sample sentences that the compiler has extracted from original Chinese texts. Discussion of examples from different types of text will probe the functioning of the quotations and the intention of the compiler. The inclusion of the quotations in the dictionary appear to be the result of both linguistic and cultural intention, in the sense that they are revealing of the syntax of Chinese but also aspects of Chinese culture. These findings will contribute to the understanding of the practice of the inclusion of literary quotations in bilingual dictionaries and the role of the compiler in introducing foreign cultural aspects.

Keywords: LITERARY QUOTATIONS, BILINGUAL DICTIONARIES, DUTCH, CHINESE, NINETEENTH CENTURY, LEXICOGRAPHICAL PRACTICE, ROLE OF THE COMPILER

Opsomming: Literêre aanhalings in tweetalige woordeboeke: 'n Gevallestudie van 'n negentiende-eeuse Nederlands–Chinese woordeboek. Die voorkoms van literêre aanhalings in woordeboeke dateer uit die dae toe die soeke na ekwivalente eerder op interteks as op intertaal gebaseer is. Vir beide eentalige en tweetalige woordeboeke was die doel van literêre aanhalings om bewys van gebruik te lewer en om die betekenis van die toepaslike trefwoord uit te lig, met dié verskil dat dit by tweetalige woordeboeke via vertaling geskied het. In hierdie artikel word dié verskynsel in 'n negentiende-eeuse Nederlands-Chinese woordeboek ondersoek deur die analisering van voorbeeldsinne wat die samesteller uit oorspronklike Chinese tekste onttrek het. In die bespreking van voorbeelde uit verskillende tekstipes sal die funksie van die aanhalings en die bedoeling van die samesteller deeglik ondersoek word. Dit wil voorkom of die insluiting van die aanhalings in die woordeboek die resultaat van beide taalkundige en kulturele doelwitte is aangesien sowel die sintaksis van Chinees as aspekte van die Chinese kultuur weerspieël word. Hierdie bevindings sal bydra tot 'n beter begrip van die praktyk van insluiting van literêre aanhalings in tweetalige woordeboeke en van die samesteller se rol in die bekendstelling van vreemde kulturele aspekte.

Sleutelwoorde: LITERÊRE AANHALINGS, TWEETALIGE WOORDEBOEKE, NEDER-
LANDS, CHINEES, NEGENTIENDE EEU, LEKSIKOGRAFIESE PRAKTYK, ROL VAN DIE
SAMESTELLER

1.   Introduction

The practice of including literary quotations in dictionaries dates from the
days when the compilation of dictionaries was based on the search for inter-
textual rather than interlingual equivalence. "In the European tradition, it was
the text that gave the initial impetus to the search for equivalence." (Adamska-
Sałaciak 2010: 388) For the well-known English lexicographer Samuel Johnson
the primary role of quotations in the *Oxford English Dictionary* (*OED*) was to
supply evidence of usage and illuminate the meaning of a word (Brewer 2010: 101).
As Brewer further notes, the quotations in the *OED* were

> supposed to be constitutive rather than illustrative of meaning: that is, the lexicog-
> raphers deduce the meaning of words from their quotations (rather than deduce
> the meanings of their quotations from their pre-determined definitions). Clearly
> these quotations and definitions are functioning quite differently, and their in-
> clusion in the *OED* looks to be the result of cultural not just linguistic intention
> on the part of the lexicographers, deriving from assumptions of the kind articu-
> lated by Burchfield, that not to consider such examples of usage 'leave one look-
> ing at a language with one's eyes partly blind-folded.' (Brewer 2010: 112-113)

Hence beyond the linguistic intention of inclusion of literary quotations in the
dictionary, there is an additional purpose: the cultural intention, i.e. the quota-
tions are included for "the purpose of educating via exposure to morally in-
structive sentiments" (Brewer 2010: 100). As a result of the inclusion of literary
sources, the *OED* ultimately functioned not only as a word-list but also as "an
intellectual history of an entire culture" (Brewer 2010: 100-101).

The aim of this study is to examine the intention of the compiler and the
functioning of the literary quotations in the nineteenth-century Dutch–Chinese
dictionary *Nederlandsch–Chineesch Woordenboek met de Transcriptie der Chineesche
Karakters in het Tsiang-Tsiu Dialekt* 荷華文語類參 *Hô Hoâ Bûn-Gí Luī-Ts'am*
[Dutch–Chinese Dictionary with Transcription of Chinese Characters in
Zhangzhou Dialect]. The compiler of this dictionary was the Dutch sinologist
Gustaaf [also spelled Gustav] Schlegel (1840–1903) who started collecting the
data for the dictionary upon his arrival in China at the end of the 1860s. During
his years in China, in the Dutch East Indies (present day Indonesia) and upon
his return in the Netherlands, Schlegel continuously worked on the dictionary
and arranged funding for publication. The publication process of the dictionary
has been described in detail by Koos (P.N.) Kuiper in his study on early Dutch
sinologists, which also contains an elaborate overview of the contents of the
dictionary and its reception (Kuiper 2017). Previously, I have examined aspects
of equivalence in the dictionary (Heijns 2020). The current article, however,

aims to analyse Schlegel's practice of including phrases that he extracted from Chinese source texts.

Schlegel considered the *Chineesch–Hollandsch woordenboek van het Emoi dialekt* [Chinese–Dutch dictionary of the Amoy Dialect] compiled by Johannes J.C. Francken (1838–1864) and Carolus F.M. de Grijs (1832–1902) as the counterpart of his dictionary, as he noted in the introduction to his dictionary (Schlegel 1886: 2). Indeed, Schlegel only produced a Dutch–Chinese dictionary and the inclusion of many Chinese concepts implies that Schlegel's work method resembled compilation of a unidirectional dictionary. Since the source language of Schlegel's dictionary is the language of the user, it can be considered an *active* or encoding dictionary, because "its purpose is to support the encoding of expressions in a foreign language" (Osswald 2015: 1981). Schlegel introduced typical Chinese cultural aspects under associated Dutch headwords. It is clear from the quotations that he had already done the translation from Chinese into Dutch beforehand and he inserted the phrases into the relevant entries in the dictionary according to terms in his Dutch version, sometimes even using the same phrase under more than one headword, where suitable.

What follows here, is an analysis of the literary quotations included in the dictionary which will shed light on the intention of Schlegel, similar to the examination of Brewer into Johnson's. The difference is of course the fact that Schlegel's is a bilingual dictionary bringing the complicating factor of anisomorphism because it is not only about the meaning and use of words but also about finding equivalents between words in two different languages. To quote Adamska-Sałaciak again: "[...], bilingual lexicography straddles the domains of linguistics and translatology" (2010: 389). Hence, in addition to discussing various examples of literary quotations and their purposes with either a linguistic or a cultural intention, I will also comment on the translation. Ultimately, I will conclude with what the inclusion of literary quotations means in terms of the functioning of the dictionary, not only as a word-list but also as an intellectual history of China's culture and Schlegel's contribution to it in terms of selection and translation.

2.   Background to the Compiler and his Dictionary

Tasked with the responsibility of teaching Chinese to Dutch candidates who passed the entrance exam to study Chinese at Leiden University, Schlegel was confronted with the lack of good teaching material. Back then the students who were admitted to the course were trained to become Chinese interpreters in the Dutch East Indies.[1] As Kuiper notes:

> In the 1870s, Schlegel had complained several times that there were no good language tools for the students. But in 1888, three of the four volumes of Schlegel's Dutch–Chinese dictionary had appeared, and the fourth volume would be completed in 1890. *This must have been a great help to the students in making translations from Dutch into Chinese.* Moreover, Francken and De Grijs's Amoy-Dutch diction-

> ary had appeared in 1882, and although it was not as good as Carstairs Douglas's dictionary, it was certainly useful. Of course, the students would also make use of Williams' and Douglas' dictionaries. (Kuiper 2017: 431, my italics)

In addition to the goal of helping translators at work in the Dutch East Indies, Kuiper emphasizes the need of dictionaries for students learning Chinese. Schlegel's teaching method was to have the students manually copy (part of) textbooks. Schlegel taught the students the spoken language of the Tsiang-tsiu [Zhangzhou] dialect, Chinese moral maxims and oral translations from Dutch into Chinese and vice versa. In Schlegel's view, reading and translating "until you think in the same vein as the Chinese" (Schlegel 1892: 48) was part of the process of learning Chinese. Study of the Chinese language prepared the students for their future career as Chinese interpreter that involved oral interpreting in court, and written translations from Dutch into Chinese and vice-versa. Having done both (study Chinese and serving in the Dutch East Indies) Schlegel was keenly aware of the urgent need of a good dictionary, even though in practice, the interpreters in the Dutch East Indies often worked with the help of Chinese clerks, by explaining orally (in Chinese) the contents of the Dutch text to be translated which the Chinese clerk would then put into written Chinese.

Schlegel's dictionary is arranged alphabetically by Dutch entries and provides Chinese equivalents in Chinese characters with Hokkien transcription. Hokkien, or *minnan* dialect, was the most commonly spoken language among ethnic Chinese in the Dutch East Indies and was therefore essential for the interpreters to learn. As Schlegel explains in the introduction, his aim was to help Dutch interpreters of Chinese in the Dutch East Indies with their work. In his idea of learning Chinese, Schlegel stressed the importance of reading and translating source texts, therefore in the dictionary, he always strove to provide (sample) sentences from original Chinese texts. Despite sporadic mention of the source of the original text, my findings show that Schlegel draws on a great number of Chinese texts, including classical texts (including the *Lunyu* 論語 [Analects], *Mengzi* 孟子 [Mencius] and *Zhuangzi* 莊子 [Zhuangzi]), historical works (*Shiji* 史記 Records of the Grand Historian, *Hanshu* 漢書 History of the Han, *Jinshu* 晉書 History of the Jin), poetry (*Shijing* 詩經 Book of Songs) and novellas (*Jingu qiguan* 今古奇觀 Wonders Old and New) as well as a mixture of other works (*Kangxi zidian* 康熙字典 Kangxi Dictionary, *Taiping guangji* 太平廣記 [Extensive Records of the Taiping Era], *Huajianji* 花箋記 Flowery Letterhead, *Wufengyin* 五鳳吟 Wufengyin). The many examples, comments and references to other experts resulted in this dictionary becoming a bulky work. It was printed in four volumes over a period of six years (1884–1890) with a total of more than 5,000 pages.

According to Schlegel, he included phrases mainly to illustrate the use of translated words: "The Chinese being emphatically a language of phrases, it was of importance to show, by extracts from native sources, the collocation in which words occur." (Schlegel 1886: 17-18) Moreover, Schlegel concurred with

Carstairs Douglas (1830–1877) and cites from the preface of his *Chinese–English Dictionary of the Vernacular or Spoken Language of Amoy*: "The whole style and character of Chinese thought and expression is so different from the nearest English (read European) equivalents, that the work of reversing a dictionary, which at first sight seems very easy, would really be enormous, falling not very far short of the original composition" (Schlegel 1886: 18; Carstairs Douglas 1873: ix).

3.   Analysis of Literary Quotations

Considering that Franken and De Grijs had opted for the vernacular in their dictionary, Schlegel himself claims that he "decided upon the written language (the so-called Book-language) for the compilation of my Dictionary" (Schlegel 1886: 2). He also stresses that "[c]omplete sentences and expressions from Chinese authors were copied out, and inserted alphabetically into the Dictionary, at the different words contained in such a sentence or expression." (Ibid.) In other words, the user can rest assured that the sample sentences were from genuinely Chinese source texts.

It is clear from what seems a random selection of quotations that Schlegel made no difference between the types of literary genres. In the pre-modern Chinese literary tradition, i.e. up until the late nineteenth-early twentieth century, literature was "intended to be useful, exert an influence on readers that was simultaneously intellectual, moral and aesthetic" (Idema and Haft 1997: 9). The genres included two literary forms, poetry and the essay, but more broadly it also included historical and philosophical works (McDougall 1971: 2). This is different from the Western literary tradition, which also included fiction and drama. Hence, given their background, some Western translators of Chinese literature with their Western ideas found that Chinese fiction provided an excellent source of information about the Chinese people and their culture, and they regarded fiction as a reflection of Chinese society (Wyllie 1964: 201-202). In China, however, it was not until the early twentieth century that norms started to change, and drama and fiction became recognized literary genres.

As stated above, Schlegel's primary aim with the dictionary was to help Dutch interpreters of Chinese in the Dutch East Indies with their work, and it was also meant to be a tool for students to learn Chinese. In terms of lexicographic functions as proposed by Henning Bergenholtz and Sven Tarp, the function of Schlegel's dictionary and the inclusion of literary quotations are primarily "communication-orientated" to solve problems "during production, reception or translation phases" of the communication process (Bergenholtz and Tarp 2003: 174). This is clear from the examples in Schlegel's dictionary as shown under "linguistic intention" below, where semantic and grammatical aspects are analysed in the quotations. Yet, another function of Schlegel's dictionary and the inclusion of literary quotations is to a certain extent also "knowledge-orientated" in that it provides "additional information on some topic, e.g. general cultural and encyclopaedic information, specialised informa-

tion regarding a scientific discipline (biology, geology etc) or information about a specific language related to the language-learning process (for example the learning of a foreign language)" (Bergenholtz and Tarp op cit.). This is clear from the examples from Schlegel's dictionary under "cultural intention" below, where additional information is given about Chinese culture and history. In the final category "failed intention" there are examples of literary quotations that are not very helpful and provide neither linguistic nor cultural help.

3.1   Linguistic Intention

A very practical example of a literary quotation from *Huajianji* can be found under the headword "aanbieden" [to offer] (Schlegel 1886: 34-35):

> *hij begon haastig bij het lamplicht eene memorie aan Z.M. [Zijne Majesteit] te schrijven, waarin hij zich aanbood om naar de grenzen te trekken en de rebellen te bestrijden.*
> [he hastily started writing a memorial to HM *[His Majesty]* by the lamplight, in which he offered to go to the border to fight the bandits.]
> 燈下連忙寫表文，願出邊關征賊寇。

The translation is accurate and the use of *yuan* 願 for "aanbieden" in the sense of "to be willing to" makes sense. In fact, this whole phrase appears useful. The description of the action (writing), the name of the document (memorial) and the topic of voluntary assistance (fighting) provide practical information and would have been useful for the translations that the translators had to do at work.

A good example of what Schlegel regards as a Chinese phrase can be found in the entry "uitlokken" [to tempt, entice] (Schlegel 1884: 312):

> *Kiun-tsan wist door het middeltje om den tijger uit het bosch te lokken, Ki-sing in zijn eigen huis te verschuilen.*
> [By means of luring the tiger out of the forest, Junzan managed to hide Qisheng in his own home.]
> 君讚用調虎離山之計，將琪生藏在自己家裡

By citing this phrase from *Wufengyin*, Schlegel shows that in Chinese "uitlokken" uses a construction with *yong* 用 [use], *diao* 調 [move] and *ji* 計 [a plan]. This is typically an example to show there is no direct equivalent, but rather a phrase that the user has to follow. The difficulty, however, might be for the user to find out the term for "uitlokken" in the Source Text (ST) because Schlegel does not give an explanation.

Another example from *Wufengyin* can be found under "Uitredden" [to rescue from] (Schlegel 1890: 322):

het was Roodbaard, die aan het hoofd van meer dan honderd bandieten Ki-sing uit de gevangenis kwam redden.
[It was Redbeard, who as the leader of more than one hundred bandits came to rescue Ki-sing from prison.]
却是紅鬚領著百餘嘍囉進來却獄救琪生。

Note how Schlegel changed the gerundive form *lingzhe* [leading] into the wordier clause "who as the leader of" which seems unnecessary. Both these literary quotations from *Wufengyin* are sample sentences with enough context for the user to understand the meaning and use.

An example of a more grammatical purpose is the suggestion by Schlegel on how to formulate questions in Chinese. Schlegel quotes from the Confucian classic the *Mencius* to give an example of an interrogative sentence in Chinese. The exact same phrase appears under the two entries "vraagwoord" (question word) (Schlegel 1890: 971) and "vragen" (to ask) (Schlegel 1890: 976):

De Koning vroeg: was het park van koning Wen niet zeventig vierkante mijlen groot?
[The King asked: wasn't the size of King Wen's park seventy square miles?]
王問曰。文王之囿方七十里，有諸。

The construction in Chinese: "clause, *is this true*?" shows the user that the last two characters *you zhu* are the question words for "is it true?". These are particles used in classical Chinese. Compare with D.C. Lau's version, which translated more literally: "Is it true that the park of King Wen was 70 li square?" (*Mencius* 2003: 29). This literary quotation is a good example of how to write interrogatory sentences in classical Chinese and the inclusion mainly functions as a grammatical instruction.

It is striking that some of the most useful and practical sentences come from the so-called *caizi-jiaren* 才子佳人 (talent and beauty) stories which are described by Cynthia J. Brokaw as:

> the "talent and beauty" (*caizi jiaren* 才子佳人) love stories, appearing first in the late Ming dynasty and continuing to be written in some numbers well into the nineteenth century. These usually rather short works concerned talented scholars and beautiful (and often even more talented) women who, after lengthy and sometimes ludicrous impediments, marry and live happily ever after. They vary widely in quality: "At their best these novels [...] are reminiscent of eighteenth-century French comedies. At their worst, however, they are pedantic and soporific Chinese equivalents of the dime novel."[2] With convoluted plot lines, most were written in a "beautiful and refined" style appropriate to their subject matter; they seem to have been intended for a fairly literate readership. (Brokaw 2007: 488-489)

Schlegel read many of these *caizi-jiaren* stories, as evidenced by the many quotations from works such as *Wufengyin*, *Huajianji* and *Yulouchun* 玉樓春 [Jade Tower Spring]. These works are written in the vernacular style, and a work

such as *Huajianji* even has some local dialect (Cantonese) included. Hence the language of these works is rather colloquial and relatively easy to read and use.

3.2   Cultural Intention

Far more difficult are the literary quotations that introduce Chinese culture. These come in the form of philosophical claims such as the following from the *Mencius* under the headword "buigen" (to bow, bend), where it is used more figuratively in the sense of "to submit, yield" (Schlegel 1886: 720).

> *die door rijkdom en aanzien niet brooddronken wordt, niet afwijkt, ook als hij arm en geminacht is, en niet voor geweld en gezag buigt, dien kan men een' groot man noemen.*
> [He] who does not become wasteful by wealth and prestige, does not stray even though he is poor and despised, and does not bow to violence and authority, he may be called a great man.
> 當貴不能淫。貧賤不能移，威武不能屈。此之謂大丈夫 (*Mencius Tengwengong xia 7*)

Written in classical Chinese, the ST is dense and compact. In order to make it work in a foreign language, it is necessary to add words when translating from classical Chinese into a European language to make it grammatically sound, because of the huge differences between the languages. This is a feature of all Chinese classical texts and not limited to Confucian texts or poems from the *Shijing* as shown below. In the example from the *Mencius*, there is also a certain pattern in the Chinese, if looked at word for word, which cannot be translated literally. In Schlegel's translation the notion of "buigen" comes out clearly, and for the three criteria of "a great man", which follows the Chinese repetition of *bu* [not] clarifies what a great man does not do. This is for instance different in comparison with James Legge's translation: "to be above the power of riches and honours to make dissipated, of poverty and mean condition to make swerve from principle, and of power and force to make bend — these charac-teristics constitute the great man" (Legge 1875: 218). Legge has omitted nega-tion, but opted for repetition of "to make + verb".

Another example is the inclusion of poetry under the headword "Rijm" [Rhyme] (Schlegel 1884: 633-634). Schlegel includes poems from the *Book of Songs* 詩經 which is a collection of over 300 poems dating from the eleventh to seventh centuries BC. Although the purpose of quoting the poems here is pri-marily to show the differences in rhyme patterns, it is also a good introduction to classical Chinese poetry. The *Book of Songs* is regarded as learning material and a source of four character-idioms known in Chinese as *chengyu*. In imperial China, every educated person would have read it and would have been familiar with the poems. For three types of rhyme patterns, there are two stanzas from the poem *Guan ju* 關雎 [Mandarin ducks], and one from the poem *Tuju* 兔罝 [Rabbit net],

while for the other six types there are only explanations, no examples. A closer look at the translations of the stanzas shows that Schlegel follows Legge's version quite closely although Schlegel is less wordy as the following stanza from *Guan ju* (Schlegel 1884: 634) shows:

> *Hij zocht haar, en vond haar niet*
> *En wakend en slapend dacht hij aan haar*
> *Hij dacht lang; hij dacht lang*
> *En wentelde zich om en om*

> [He sought her and found her not
> And waking and sleeping he thought of her
> He thought long; he thought long
> And turned around and around]

> 求之不得
> 寤寐思服
> 悠哉悠哉
> 輾轉反側

From the ST we can see that it is four characters per line, a pattern which is lost in translation. The ST is extremely terse, for example the four characters of the first line word-for-word is literally: seek-it-not-obtain. When compared with Legge's version, the first two lines are translated in a similar way: "He sought her and found her not, / And waking and sleeping he thought about her" (Legge 1871, as quoted in Minford and Lau 2002: 75). But then the third and fourth lines, where Legge has: "Long he thought; oh! long and anxiously / On his side, on his back, he turned and back again" (Ibid.).[3] This shows that Legge's version is wordier than Schlegel's, who sticks more closely to the ST, retaining the repetition in line 3 just like in the Chinese, and in line four he retains the meaning of the ST where Legge is overtranslating: adding "anxiously" in the third line and an extra turn in the fourth line. Overall, the inclusion of literary quotations from the *Book of Songs* is informative about Chinese poetry, even though not directly practical or useful for the user's translation tasks. However, there was a poetry requirement for the major period of the official imperial examinations *keju* in China, in connection with the notion that "the study and practice of poetry encouraged careful writing" (Su Shi's opinion according to Murck 2000: 52). It is likely that, in Schlegel's view, inclusion of poems and reading poetry would help the student or translator of Chinese, not necessarily in the practice of the language but rather the understanding of it as "culture". So in all probability, besides showing the rhyme patterns, for Schlegel it must have been a method of exposing the user to Chinese poetry, purely for poetry's place in the Chinese literary tradition.

An even more culturally loaded entry is the headword "Fata-Morgana", with "Luchtspiegeling" in brackets, in the sense of "mirage" or "optical illusion." Here Schlegel quotes from and refers to a total of six different Chinese and

Western texts as evidence for the meaning and use of the word, hence the entry reads almost like a research note. (Schlegel 1886: 1168-1170) After giving some basic equivalents, including *huanjing* 幻境, *shen* 蜃, *shenlou* 蜃樓 etc., Schlegel cites the description from the entry *shen* 蜃 [clam] in the Chinese reference work the *Kangxi Dictionary*. The description explains the origin of the clam in connection with the phenomenon of "mirage", that is that *shen* is "a kind of dragon whose exhaled breath forms the shape of towers and cities. If it starts raining one can see them and one calls them *Shen* towers or Seamarket." This is a kind of optical illusion. Here, Schlegel refers to the French engineer and sinologist Édouard Biot's (1803–1850) 'Note sûr un phénomène de mirage Indiqué par quelques textes chinois' [A note on the phenomenon of "mirage" as indicated by some Chinese texts] published in *Journal Asiatique* 1848/07: 518-520. Here, Schlegel adds in brackets that "the text has been translated inaccurately [by Biot]" as a warning to the user (Schlegel 1886: 1168-1169).

Among the most important quotations in this entry that enlighten the user are from the Taoist work *Zhuangzi* and historical works *History of the Jin* 晉書 and *History of the Song* 宋書, that both convey the phenomenon of mirage in cultural sense, but will be difficult for the user to apply in translation. Schlegel notes that according to Zhuangzi "wild horses and dust" 野馬也塵埃也 were mirages. Schlegel goes on to explain this concept with quotes by commentators on the *Zhuangzi*, among which is the following quotation (Schlegel 1886: 1169):

> *De commentator op de "Wilde paarden" van Zhuangzi zegt: wanneer de zon op het stof schijnt, en een lichte wind die in de wildernis doet stuiven, noemt men dit, met een woordverdraaiing, Zonnegloed; de eenvoudige man, die dit verschijnsel ontwaart, noemt het "Wilde paarden"; de dorstige, die het ontwaart, houdt het voor stromend water.*
> [The commentator on "Wild horses" of Zhuangzi says: when the sun shines on dust and a light wind blows it into the wilderness, this is called "solar glow" in a word distortion; the simple man who observes this phenomenon calls it "wild horses"; a thirsty person who perceives it sees it as running water.]
> 莊子野馬注云。日光著塵，微風吹之曠野中。轉名之為陽燄。愚夫見之，謂之野馬。渴人見之，以為流水。

In terms of language, these sentences in Chinese are again terse and pithy. In this quotation, there is enough context to understand the notion of "wilde paarden" [wild horses]. The motivation to include this phrase must have been the explanation of the idea of "wild horses" and how it is interpreted. Problematic in the translation is the meaning of *zhuan ming* which Schlegel translates as "woordverdraaiing" [word distortion]. Incorrect punctuation has separated *zhong* from *zhuan*, meaning to shift or transfer. The action of the wind causes the dust to move and that changed its name into "solar glow." On the whole, the quotation is illuminating of the concept of "wild horses" but it is not very convincing in terms of being equivalent to "fata-morgana".

In terms of meaning, the literary quotations from other historical sources come closer to the idea of a "mirage" in the sense of "optical illusion". Two examples will illustrate this. The first "Water reflection" 水影 is from the *History of the Jin* (Schlegel 1886: 1169) and the second "Earth mirror" 地鏡 is from the *History of the Song* (Schlegel 1886: 1169-1170):

> *Ten tijde van Fu Jian (AD 357–385) zag men in Chang'an eene Waterspiege-ling, die van verre er uitzag als water; zag men op de aarde, dan ontwaarde men echter menschen. Zij duurde tot het jaar van Jian (385) en verdween toen.*
> [During the Fu Jian (AD 357–385) era the people in Chang'an saw a Water reflection which looked like water from afar; but on the earth they saw people. This continued until the year of Jian (385) and then dis-appeared.]
> 符堅時長安有水影。遠觀若水。視地則見人。至堅晚年而止
>
> *in het 25e jaar der periode Yuanjia der regeering van Keizer Wen der Song-dynas-tie (AD 448), zag men des winters verre ten zuiden der stad Qingzhou iets als de schaduw van water in het land, hetgeen men "Aardspiegeling" noemde.*
> [In the twenty-fifth year of the Yuanjia administration ruled by Emperor Wen of the (Liu) Song Dynasty (AD 448), people saw in winter far south of the city of Qingzhou a kind of shadow of water in the ground, which they called "Earth mirror".]
> 宋文帝元嘉二十五年，冬，青州城南遠觀見地中如水有影，謂之地鏡

The factual description in Chinese is rewritten by Schlegel in a more descrip-tive way in Dutch which is suitable for history writing. Problems in the trans-lation are: in the first citation, *wannian* is not necessarily just only that one final year of 385, it is more in general "the final years" of Jian's reign period. In the second citation, his translation of "*ying*" as "shadow" is problematic, it should be reflection, which then corresponds with what you see in the mirror. The additional information about which the Chinese lunar year corresponds with the Western (Gregorian) calendar is very helpful. Both quotations describe the phenomenon more closely to what is understood as "mirage". It appears that Schlegel's attempt to include the Taoist concept of wild horses under "mirage" is somewhat contrived, but it was supposedly the closest concept that he could find. For the meaning of "mirage", the historical sources are more suitable. Nevertheless, it is clear from the quotes and examples under headword "fata morgana" (mirage) that Schlegel studied the "mirage" issue thoroughly, giving thought to its meaning and trying to find an equivalent. The context of the use of "wild horses and dust", "water reflection" and "earth mirror" is clear and helpful. Moreover, the many examples from historical sources show that in dif-ferent periods and different places mirages were observed in China.

3.3   Failed Intention

There are literary quotations that are difficult to categorize. They seem to have

neither linguistic nor cultural purpose. Take for example this quotation from "The Oil Vendor and the Queen of Flowers" 賣油郎獨佔花魁女 (hereinafter "The Oil Vendor") under "Bespatten" [splash, spatter] in the dictionary (Schlegel 1886: 452):

> *ik reken mij gelukkig, dat zij (mijne kleeren) door eenige droppels vocht van Mejuffrouw bespat zijn*
> [I considered myself lucky that they (my clothes) were splashed by a few drops of moisture from the young lady.]
> 有幸得沾小娘予之餘瀝

At the beginning of the entry, Schlegel provides the more commonly used character *jian* 漸 for this term. Although the combination of *"zhan* 沾*... li* 瀝*"* has the same meaning here, the entire sample sentence is quite unclear. There are too many questions: why is he feeling lucky? What type of liquid is the young lady splashing? What kind of situation would this happen? The only clue Schlegel gives is "my clothes" in brackets. It appears that there is no practical purpose for the user, neither in lexical nor in grammatical sense and the line is quoted out of context. "The Oil Vendor" is a story that Schlegel had previously published in French as "Le Vendeur-d'huile qui seul possède la reine-de-beauté." As Kuiper notes: "This [publication] became one of his textbooks when teaching in Leiden, and it was read by all his students." (Kuiper 2017: 131) It is a story from *Stories Old and New* 今古奇觀 a collection of forty novellas originally published anonymously, but whose author was later identified as Feng Menglong 馮夢龍 (1574–1646).

Another dubious quotation is from the text *Yulouchun* by Baiyun Daoren [White Cloud Daoist], and concerns the explanation of an allusion in the entry *bruidsnacht* [wedding night] (Schlegel 1886: 715):

> *Van dit enkele bezoek naar den Yang toren in den bruidsnacht, ben ik dadelijk met een droom van een beer begunstigd, en ben nu reeds een half jaar zwanger.*
> [After this single visit to the Yang tower on my wedding night, I was immediately favoured with the dream of a bear and now I have been pregnant for six months.]
> 妾身花燭之夜一赴陽臺，遂符熊夢。今已懷娠半載

If read literally, the first part of this quotation is rather cryptic: "my body in the night of candles with dragon and phoenix patterns went to the Yang tower, then had the dream of a bear". "The night of candles" stands for "wedding night," and "dream of a bear" implies that one will give birth to a son. By giving this example, Schlegel shows the use of symbols and metaphors in Chinese texts. This may be interesting information for those reading Chinese literature, but not necessarily practical for the user.

The examples of literary quotations discussed here show the great variety of texts that Schlegel had at his disposal. Even without indication of the source or explanation of the nature of the texts, many of these sample sentences are

useful for the user. Some are practical in the sense that the meaning and use are clear and ready for use, some are interesting in that they provide information about Chinese culture. However, there are also quotations that provide neither information nor practical use and are quoted out of context. Overall, the user is exposed to a wide variety of genres of texts and different types and register of Chinese language. This kind of material is especially useful for teaching Chinese language and literature.

4.   Influential Thinking

It is clear from the data incorporated in Schlegel's dictionary that not all correspond with its lexicographic functions as identified above. As Bergenholtz and Tarp write: "no data whatsoever should be included in a dictionary if it cannot be argued on the basis of its respective functions" (Bergenholtz and Tarp 2003: 177). Schlegel's failure to do so must have in part been the result of the limited availability of resources and perhaps also a lack of awareness. However, the selection of quotations and the method of translation must have also been influenced by his personal ideas and interests, in spite of the fact that the general expectation is that the data in the dictionary are objective. As Atkins and Rundell note:

> Lexicographers, like historians, are expected to be "neutral" recorders of facts—but this is not as straightforward as it sounds (in either discipline). Departures from lexicographic neutrality are characterized as "editorializing", which is seen as a reprehensible tendency. (Atkins and Rundell 2008: 427)

This so-called "editorializing" or translational intervention is manifest in Schlegel's comments, explanations, and quotations in the dictionary. It appears that Schlegel was not alone in working this way, however. Atkins and Rundell cite the example of the aforementioned lexicographer Samuel Johnson, who, according to them, used the dictionary "to pursue personal vendettas" (Atkins and Rundell 2008: 428). In Schlegel's case, I would argue that he used the dictionary "to pursue personal interest," resulting in entries that clearly depart from lexicographic neutrality.

There is a direct link between the dictionary and the compiler's research field and literary preferences. Topics that Schlegel researched and wrote about include Chinese triads (*Thian Ti Hwui: The Hung-League Heaven–Earth*), prostitutes (*About Prostitution in China* in 1866), and Chinese astronomy (*Sing Chin Khao Youen* 星辰考原 *Uranographie Chinoise ou preuves directes que l′astonomie est primitive et originaire de la Chine* in 1875). Literary works that he had translated and published include: the *Huajianji* translated into Dutch as *De geschiedenis van het gebloemd briefpaper* [History of the Flowered Letterhead] in 1866 and aforementioned French translation *Le Vendeur-d′huile qui seul possède la reine-de-beauté ou splendeurs et misères des courtisanes chinoises.* [The Oil Vendor and the Courtesan, or the Splendors and Miseries of Chinese Courtesans] in 1877.

Furthermore, the method of translation is influenced by Schlegel's idea

that Chinese is a language of phrases as explained above, he also thought that the Chinese way of thinking differed from the Dutch. He explains this in the entry "troost" [consolation] as a noun (1890: 215):

> *zie verder "troosten", daar een abstractum "Troost", als persoonlijkheid, niet in de Chinese taal kan gedacht worden, maar steeds door het verbum uitgedrukt wordt; b.v.: "dit is mijn eenige troost" wordt vertaald met "dit is het eenige, wat mij troost."*
> [See "troosten", since an abstractum "Consolation" as a personality cannot be perceived in the Chinese language, but is always expressed through a verb, for example: "this is my only consolation" is translated into "this is the only thing that comforts me."]

According to Schlegel, "troost" cannot be used as a noun in Chinese. It only works as a verb. (This is quite the opposite of the example above of the gerundive *lingzhe* which he changed into a noun phrase.) Therefore, he refers to the entry on the verb "troosten" [to comfort]. This is not because of a linguistic difference but a matter of differences in the way of thinking. A similar case can be found in the long entry "Recht" (justice, law, right) (Schlegel 1884: 577-580), where he explains that "the notion of 'recht' is not so clear to the Chinese mind as to the Western mind, which transforms abstractions into concreta as a result of his mythological education."

It is beyond doubt that Schlegel considered his target user groups: his students in Leiden and his prospective users in the Dutch East Indies — after all, he was once a student and interpreter too, and he was aware of their needs. But of course, the information he provided was largely dependent on his own knowledge, interests, ideas and the availability of material.

5.   Conclusion

Schlegel cited original Chinese texts to explain the syntactical properties of words and to illustrate authentic use of Chinese words. In many cases the inclusion of literary quotations helps the user to understand the meaning and use of the relevant word. The translation provided by Schlegel is mostly accurate and helpful to the user. The abundant choice in various literary works shows that Schlegel was well-read. Judging from the entries, it appears that he quoted randomly from different types of literary works which are written in a variety of levels and different registers of the Chinese language, ranging from the colloquial and dialect to formal and philosophical. These differences do not show in translation and it is questionable how the user would be able to determine suitability. Another problem is quoting out of context, in some cases it is difficult to understand the sentences on their own.

In spite of the fact that Schlegel was very much aware of his user situation and he set himself an aim of helping translators and students, and so in terms of lexicographic functions, the inclusion of literary quotations in the dictionary

is primarily communication-orientated but to a certain extent also knowledge-orientated, in that it provides more than linguistic help. Nevertheless, it seems that Schlegel did not stick to his own "rules" and it seems more likely that he just used data that was available to him. Hence, in this context, the idea that Brewer puts forward that "Lexicographers deduce the meanings of words from their quotations (rather than deduce the meanings of their quotations from their pre-determined definitions)" is true also of Schlegel. That is in addition to the idea that words in quotations and their original definition function differently. Thus says Brewer, in the case of *OED* the inclusion of those quotes seems to be the result of cultural and not just linguistic intention on the part of the lexicographer. In the case of Schlegel, it is the result of cultural, linguistic and phraseological intention of the compiler.

Finally, the fact that Schlegel cites from a large variety of texts, not only the ones that count as "Chinese literary tradition" but also those that are considered "trivial literature" shows that his interest was quite broad. By including all the important works that make up the greater part of Chinese literary tradition, the dictionary maybe considered an anthology of Chinese literature. It is striking though that the more practical sample sentences derive from trivial literature, and seem to have a linguistic intention, whereas the culturally loaded sentences are quoted from works regarded as Chinese literature in the literary tradition and enlighten the user on important aspects of Chinese culture. Although this study is merely an interpretation of literary quotations included in an existing, very old, dictionary, hopefully the findings will contribute to the future development of lexicographic theories and practices. The issue of including literary quotations in the dictionary is in itself very important, and worth contemplation by prospective dictionary compilers. In the case of deciding to include literary quotations, hopefully the examples in this study will help shed light on their criteria and suitability.

Acknowledgement

Endnotes

1.   Kuiper 2017 for details on the training.
2.   The full quote is: "At their best these novels, with their frequent disguises and misunderstandings, are reminiscent of eighteenth-century French comedies. At their worst, however, they are pedantic and soporific Chinese equivalents of the dime novel." Idema and Haft 1997: 227.

3.   There is also a versified version by Legge: Sought for her long, but all his search was vain./
Awake, asleep, he ever felt the pain/ of longing thought, as when on restless bed, / Tossing
about, one turns his fevered head. (Minford and Lau 2002: 77).

References

A.   Dictionaries

Douglas, Carstairs. 1873. *Chinese–English Dictionary of the Vernacular or Spoken Language of Amoy.*
London: Trübner & Co.

Francken, J.J.C. and C.F.M. de Grijs. 1882. *Chineesch–Hollandsch woordenboek van het Emoi dialekt*
[Chinese–Dutch Dictionary of the Xiamen Dialect]. Batavia: Bataviaasch Genootschap van
Kunsten en Wetenschappen.

Schlegel, Gustaaf. 1884. *Nederlandsch–Chineesch Woordenboek met de Transcriptie der Chineesche
Karakters in het Tsiang-Tsiu Dialekt.* 荷華文語類參 *Hô Hoâ Bûn-Gí Luī-Ts'am* [Dutch–Chinese Dic-
tionary with Transcription of Chinese Characters in Zhangzhou Dialect]. *Volume 3 (O–S).*
Leiden: Brill.

Schlegel, Gustaaf. 1886. *Nederlandsch–Chineesch Woordenboek met de Transcriptie der Chineesche
Karakters in het Tsiang-Tsiu Dialekt.* 荷華文語類參 *Hô Hoâ Bûn-Gí Luī-Ts'am* [Dutch–Chinese Dic-
tionary with Transcription of Chinese Characters in Zhangzhou Dialect]. *Volume 1 (A–G).*
Leiden: Brill.

Schlegel, Gustaaf. 1887. *Nederlandsch–Chineesch Woordenboek met de Transcriptie der Chineesche
Karakters in het Tsiang-Tsiu Dialekt.* 荷華文語類參 *Hô Hoâ Bûn-Gí Luī-Ts'am* [Dutch–Chinese Dic-
tionary with Transcription of Chinese Characters in Zhangzhou Dialect]. *Volume 2 (H–N).*
Leiden: Brill.

Schlegel, Gustaaf. 1890. *Nederlandsch–Chineesch Woordenboek met de Transcriptie der Chineesche
Karakters in het Tsiang-Tsiu Dialekt.* 荷華文語類參 *Hô Hoâ Bûn-Gí Luī-Ts'am* [Dutch–Chinese Dic-
tionary with Transcription of Chinese Characters in Zhangzhou Dialect]. *Volume 4 (T–Z).*
Leiden: Brill.

B.   Other literature

Adamska-Sałaciak, Arleta. 2010. Examining Equivalence. *International Journal of Lexicography* 23(4):
387-409.

Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography.* Oxford/
New York: Oxford University Press.

Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H. E. Wiegand's Recent
Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.

Brewer, Charlotte. 2010. The Use of Literary Quotations in the *Oxford English Dictionary*. *The Review
of English Studies* (New Series) 61(248): 93-125.

Brokaw, Cynthia J. 2007. *Commerce in Culture: The Sibao Book Trade in the Qing and Republican Periods.*
Cambridge, Mass.: Harvard University Press.

Heijns, Audrey. 2020. In Pursuit of Genuine Chinese Equivalents: The Case of a Nineteenth-Century Dutch–Chinese Dictionary. *International Journal of Lexicography* 33(1): 104-115.

Idema, Wilt and Lloyd Haft. 1997. *A Guide to Chinese Literature.* Ann Arbor: Center for Chinese Studies, University of Michigan.

Kuiper, Koos (P.N.). 2017. *The Early Dutch Sinologists (1854–1900): Training in Holland and China, Functions in the Netherlands Indies.* Leiden: Brill.

Legge, James. 1875. *The Chinese Classics. Volume 2: The Works of Mencius.* London: Trübner.

McDougall, Bonnie S. 1971. *The Introduction of Western Literary Theories into Modern China, 1919–1925.* Tokyo: Centre for East Asian Cultural Studies.

Mencius. 2003. Translated by D.C. Lau. Hong Kong: Chinese University Press.

Minford, John and Joseph S.M. Lau (Eds.). 2002. *Classical Chinese Literature. Volume I: From Antiquity to the Tang Dynasty.* New York: Columbia University Press / Hong Kong: Chinese University Press.

Murck, Alfreda. 2000. *Poetry and Painting in Song China: The Subtle Art of Dissent.* Cambridge, Massachusetts/London: Harvard University Asia Center for the Harvard-Yenching Institute.

Osswald, Rainer. 2015. Syntax and Lexicography. Kiss, Tibor and Artemis Alexiadou. 2015. *Syntax — Theory and Analysis. An International Handbook. Volume 3: Syntax and Lexicography:* 1963-2001. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK) 42(3). Berlin: De Gruyter Mouton.

Schlegel, Gustaaf. 1892. *La Stèle Funéraire du Téghin Giogh, et ses Copistes et Traducteurs Chinois, Russes, et Allemands* [The Funerary Stele of Téghin Giogh, and its Copiers and Chinese, Russian and German Translators.] Leiden: Brill.

Wyllie, Alexander. 1964. (c1867.) *Notes on Chinese Literature.* New York: Paragon Book Reprint Corp.

Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvement

Fang Huang, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China (huangfang@gdufs.edu.cn)*
and
Sven Tarp, *Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; Department of Afrikaans and Dutch, Stellenbosch University, South Africa; 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: Digital applications to assist language learning are becoming increasingly popular. They typically incorporate one or two dictionaries to improve the service so that users avoid leaving the app to consult external resources. This paper deals with the two dictionaries used in a learning app for Chinese learners of English. Initially, it describes the functioning of the app as well as the two dictionaries that have different roles in the app. It then focuses on the one that is integrated into the course texts and can be activated by clicking on a word or a multiword unit. A number of deficiencies are discussed such as inconsistent treatment of words and senses, data overload, difficult access, and inconvenient location of the pop-up window that displays the lexicographical items. These deficiencies may impact negatively on the learners' motivation and the learning process in general. The paper traces the problems to the database that sustains the dictionary as well as the design of the user interfaces that filter the data offered to the users. Finally, and inspired by the classical Chinese Xun Gu tradition, it suggests an alternative, context-adapted approach that breaks with traditional features of the dictionary article and reduces the content of the pop-up window to an absolute minimum. The idea is to avoid a consultation process that interrupts the learners' reading flow and focus on learning.

Keywords: MEDIA CONVERGENCE AGE, ENGLISH LEARNING APPS, INTEGRATED DICTIONARIES, EMBEDDED DICTIONARIES, LEXICOGRAPHICAL DATABASES, USER INTERFACES, POP-UP WINDOWS, CONTEXT-AWARENESS, XUN GU TRADITION, PERSONALIZATION, INCIDENTAL AND INTENTIONAL LEARNING

Opsomming: Woordeboeke geïntegreer in Engelse aanleerderstoepassings: Kritiese kommentaar en voorstelle ter verbetering. Digitale toepassings wat as hulpmiddels in die aanleer van taal gebruik word, word toenemend gewild. Kenmerkend van hierdie digitale toepassings is die inkorporering van een of twee woordeboeke om die diens wat gelewer

word, te verbeter sodat gebruikers nie die toepassing hoef te verlaat om eksterne bronne te raadpleeg nie. In hierdie artikel word aandag geskenk aan die twee woordeboeke wat in 'n aanleerderstoepassing vir Chinese leerders van Engels gebruik word. Eers word 'n beskrywing gegee van die funksionering van sowel die toepassing as van die twee woordeboeke wat verskillende funksies in die toepassing vervul. Daarna word gefokus op die woordeboek wat in die kursushandleidings geïntegreer is en wat geaktiveer kan word deur op 'n woord of meerwoordige eenheid te klik. 'n Aantal leemtes soos die inkonsekwente hantering van woorde en begrippe, data-oorlading, moeilike toegang, en hinderlike posisie van die opwipvenster wat die leksikografiese items vertoon, word bespreek. Hierdie gebreke mag 'n negatiewe effek op die motivering van die aanleerders en die aanleerproses in die algemeen hê. Die probleme word teruggevoer na sowel die databasis wat die woordeboek steun as na die ontwerp van die gebruikerskoppelvlakke wat die data wat aan die gebruikers aangebied word, sif. Laastens, en geïnspireer deur die klassieke Chinese Xun Gu-tradisie, word 'n alternatiewe, konsep-aangepaste benadering wat afwyk van die tradisionele kenmerke van die woordeboekartikel en wat die inhoud van die opwipvenster tot 'n absolute minimum beperk, voorgestel. Sodoende word 'n konsultasieproses wat die aanleerders se leesvloei onderbreek, vermy en word daar gefokus op die aanleerproses.

Sleutelwoorde: ERA VAN MEDIAKONVERGENSIE, ENGELSE AANLEERDERSTOEPAS-SINGS, GEÏNTEGREERDE WOORDEBOEKE, INGEBEDDE WOORDEBOEKE, LEKSIKOGRA-FIESE DATABASISSE, GEBRUIKERSKOPPELVLAKKE, OPWIPVENSTERS, KONTEKSBEWUST-HEID, XUN GU-TRADISIE, VERPERSOONLIKING, TOEVALLIGE EN DOELBEWUSTE LEER

1.   Introduction

Lew (2007: 212) observed that lexicography "grows, branches out, and special-izes". This holds true not only for the highly varied topics treated by the millennial discipline but also for the many new ways of presenting the final product to its end users. If it was the case in 2007, it is even more so today with the digital revolution developing almost exponentially. There are still countries, where some dictionaries are printed in relatively big editions. But the general tendency, both here and elsewhere, is the increasing publication of lexicographical products on a big variety of digital platforms. In China, the current transition to these platforms is referred to as *the media convergence age*, a concept that also includes videos, illustrations, interaction, and other ways of meeting user needs; see Zhang (2019a, b), Li and Wang (2020), and Kang (2020). According to Fuertes-Olivera et al. (2018: 156-57), the overall transition can be further specified as a triple transformation from "the traditional stand-alone dictionary, either printed or online, to the integrated dictionary"; from "the standardized dictionary (…) to a more personalized dictionary"; and from "the dictionary as such to lexicographical data for different uses".

The first of these three tendencies is obvious. The past few years have seen the integration of dictionaries into digital products like e-readers, writing assistants, translation tools, and learning apps; see Bothma and Prinsloo (2013), Tarp et al. (2017), and Ma (2019), among others. However, as we will show in

this contribution, it seems to be much more difficult to design a lexicographical product that provides personalized assistance to its users. This has probably to do with deep-rooted traditions combined with a certain degree of resistance to sacrificing traditional dictionary features to the need for a more intuitive, digital-era design that focuses on the specific set of lexicographical data required by the users in each situation; that is, the third transformation referred to by Fuertes-Olivera et al. (2018).

In China, there is a millennial tradition called *Xun Gu* that may inspire us. When the old scribes copied manuscript works from earlier periods they frequently inserted glosses into the text to explain difficult or obsolete words and interpret the content; see Yong and Peng (2008: 21-25). The tradition continues even today when classical Chinese works are re-published. A similar tradition can be found in other parts of the world. McArthur (1986), Stathi (2006), and Hanks (2013) report that it was practiced in Europe in the Classical Greek Period, where the glosses subsequently were compiled into glossaries. This gave birth to the traditional dictionary form as we have known it ever since; that is, the dictionary as a structured collection of articles containing glosses (lemmata) and additional, explanatory items of various types.

However, the glossaries did not mark the beginning of lexicography as such, as it was already practiced embryonically by the scribes when they inserted glosses into manuscript copies. Tarp and Gouws (2019) characterize this practice as "lexicographical contextualization and personalization". In the classical Greek and Chinese traditions, the response to user needs is provided directly in the context where these needs occur, the consultation is intuitive, and the search time close to zero. In addition, the response consists exclusively of the lexicographical data required to meet the user's needs in a concrete situation, thus avoiding the adverse phenomenon of "information anxiety" (Wurman 1990). By contrast, data overload in lexicographical products may lead to a time-consuming consultation with the risk of retrieving the wrong information or no information at all, as discussed by Gouws and Tarp (2017). Hence, the *Xun Gu* and similar experiences seem to be a good starting point for a more personalized — and contextualized — lexicographical service in integrated digital products like the ones mentioned above.

Learning apps have already been discussed from different angles in the academic literature; see Zhang and Jin (2017), Rosell-Aguilar (2018), and Loewen et al. (2019), among others. In this contribution, we will focus on dictionaries integrated into second-language learning apps and approach them from the perspective outlined above. To our knowledge, this topic has not yet been treated comprehensively in the lexicographical literature. As a case study, we have chosen the *Kaiyan OpenLanguage* app that is used by Chinese learners of English as a second language. In the next section, we will argue why we have chosen this app among other possible candidates and explain how it works. In Section 3, we will describe the two dictionaries used in the *Kaiyan OpenLanguage* app. In the subsequent section, we will analyze how they perform and to what degree the design features satisfy the needs of Chinese learn-

ers. Section 5 will summarize the effect on users and consultation processes, whereas Section 6 will discuss and suggest various ways of improving the app from a lexicographical perspective. Finally, we will conclude that most of these suggestions can be generalized and applied in other learning apps.

2.   Description of English learning apps

With an overwhelming majority of Chinese people owning a smartphone or tablet, L2 learning apps, especially English ones, are becoming more accessible and popular in China. Among the learning apps can be mentioned *Shanbei* (扇贝), *Baicizhan* (百词斩), *Yingyu Qupeiyin* (英语趣配音), *Yingyu Tingli* (英语听力), and *Kaiyan Openlanguage* (开言英语). Some of these apps, like *Shanbei* and *Baicizhan*, allow most of their functionalities to be used freely. Others, like *Kaiyan OpenLanguage*, only offer a few free trials, after which their users have to buy a premium version to get access to most of the content and functionalities. Compared to traditional classroom teaching with time and location constraints, learning apps allow learners with smartphones or tablets, as well as Internet access, to attend online English classes at any time, at almost any location. They can stop when they are busy and continue learning when they are free. They can make plans for themselves and progress at their own pace. Learning apps can, therefore, be designed to cater for the users' individual needs.

When learners use L2 learning apps, the integrated dictionaries are indispensable. Consultation of the meaning or grammar of unknown words helps them better understand the context and the whole content of the course they are following. In fact, these learning apps always incorporate one or two dictionaries. When studying the course material, learners frequently encounter difficulties related to words or word forms, which they either don't know, are uncertain about, or just want to confirm. In such cases, they can immediately resort to the integrated dictionary by simply touching or clicking on the word they want to consult. A box with a dictionary article (or an extract of such an article) will then appear directly above or below the activated word and allow the learners to retrieve the required information in terms of meaning or grammar to better understand the text.

Integrated dictionaries make up a new type of digital dictionaries. They are handier and more practical than the common dictionary apps because their users do not need to close the running learning app and start a separate dictionary app to look up words for more details. In this way, learners can easily and timely get access to the relevant lexicographical data. By reducing the time spent on consultation, they can focus more on the workflow and learning process without being interrupted by the constant switch between different apps. Thus, the whole learning process can be more efficient with the help of integrated dictionaries.

Kaiyan OpenLanguage is a very popular English learning app in China. It incorporates a podcast named *Panji and Jenny tell you* (潘吉Jenny告诉你), which

is also commercialized separately and ranks among the top popular English learning podcasts for Chinese learners in the App Store. *Kaiyan OpenLanguage* offers audio courses, vocabulary courses, pronunciation courses, and reading courses, among others. It is under continuous development with several newly added interactive video courses.

The main *Kaiyan OpenLanguage* course consists of a large number of virtual classes divided into different levels from A1 to C2. Each class has four parts: *vocabulary*, *lesson* (texts in dialogue or video clips), *exercise* (on vocabulary, translation, reading, and listening), and *oral practice* (imitation). *Vocabulary* is a list of new words that are explained in the lesson. The *lesson* is composed of a text mainly in the form of dialogues. Two teachers (one native English and one Chinese English teacher for beginners and two natives for intermediate levels) will teach the lesson together by explaining new words, grammar, content, and cultural issues occurring in the text. Their way of teaching is very relaxing and interactive, like two friends chatting with each other. Some newly uploaded lessons are in the form of video clips. The *exercise* consists of multiple choices, match pairs, and fill in blanks to test if users have mastered the lesson, to practice their vocabulary, translation, reading, and listening. *Oral practice* is mostly given in form of a dialogue between two or more people from the lesson, where the learners have to play one of the roles. After the oral practice, learners will receive a score for their performance that indicates whether their pronunciation is good or should be improved.

The class does not last long. Each lesson is about 10 to 30 minutes, with exercises and practices taking another 20 minutes. It also has three complementary parts: grammar point, practical sentences, and culture, to enrich and deepen the understanding and learning of the lessons. It is a very comprehensive and practical English learning course for Chinese learners, as it is explained on its website:

> Our courses involve real-world scenarios such as traveling, workplace, food culture, including grammar, dialogue, vocabulary, and other modules to learn, making it easy for you to use the fragmented time to learn spoken English that you can use right away.[1]

Unlike the *Kaiyan OpenLanguage* app, other learning apps such as *Baicizhan*, *Yingyu Tingli*, and *Shanbei* only focus on one aspect of the English learning process. For instance, *Baicizhan* (百词斩) and *Shanbei Danci* (扇贝单词) focus on vocabulary, whereas *Baicizhan AiYuedu* (百词斩爱阅读) and *Shanbei Yuedu* (扇贝阅读) emphasize reading, *Yingyu Tingli* listening, and *Yingyu Qupeiyin* oral practice. Although some companies offer various learning services, they commercialize these services in separate apps. Baicizhan company, for instance, markets vocabulary and reading in two separate learning apps: *Baicizhan* and *Baicizhan Yuedu*. This is also the main reason why we have chosen the *Kaiyan OpenLanguage* to conduct a case study on the design features and performance of integrated dictionaries.

3.   Presentation of the integrated dictionaries

Two dictionaries are used in the *Kaiyan OpenLanguage* app: *Concise Dictionary* and *Oxford English–Chinese Dictionary*. No further information is provided about publishing houses, specific editions, or updates. The publisher of the second dictionary is undoubtedly Oxford University Press (maybe in association with a Chinese partner), and we suppose that the Concise dictionary is published by the same company as the learning app, although it is not confirmed anywhere. The two dictionaries have different status and can be accessed in different ways.

During all learning activities, except for those related to the exercise part, learners can touch or click on a word to get lexicographical assistance. A black box will immediately pop up (see Figure 1). This pop-up window may contain different types of lexicographical data such as pronunciation, part of speech, variants, senses, definitions, equivalents, and even cultural data. Users can also look up words that do not appear in the canonical form; for instance, verbs in present participle or past tense (*implementing*, *thought*), and nouns in plural (*bosses*, *trips*). They will then be connected directly to the canonical word form with a short explanation of the relation between the canonical form and the one occurring in the text. When the users, for example, click on the word *thought* in the text, the lemma *thought* will appear in the pop-up window together with various senses structured according to part of speech. In the second line, *thought* is classified as a verb with a gloss indicating that it is the past tense of *think*. If learners are not satisfied with the default data presented in the pop-up window, they can click on the signifier 查看完整释义 at the bottom to access the whole dictionary article with more lexicographical items (see Figure 1). The default dictionary used in this type of consultation is the *Concise Dictionary* that contains both a bilingual English–Chinese part and a monolingual English one. However, if this dictionary does not contain the lemma in question, the user can have a second try in the *Oxford English–Chinese Dictionary* as will be explained below. In this respect, the latter seems to function as a safety net to guarantee a better user service.
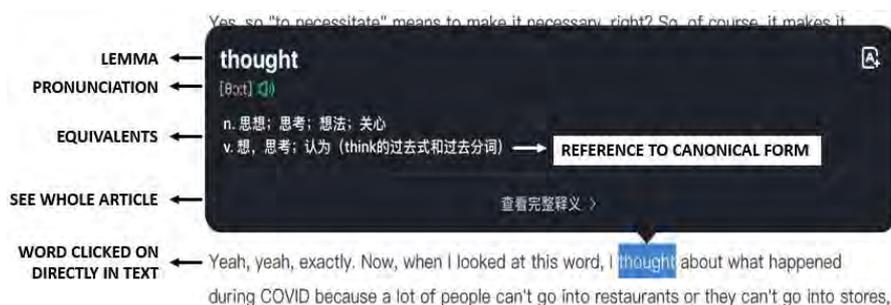
Figure 1:   Lexicographical pop-up window activated from the text

Apart from the described form of consultation, the *Kaiyan OpenLanguage* app also makes allowance for another way of looking up in the dictionaries. At the bottom of the front page, the app has a search box with the signifier 查词 (Consult a word), which affords direct access to the *Concise Dictionary*. A pop-up window will then be displayed with a traditional dictionary article that provides definitions in English and concise explanations in Chinese, as well as a list of lessons where the word appears. Users who prefer to consult the *Oxford English–Chinese Dictionary* can do so by clicking on the signifier 牛津 (Oxford) that provides access to this dictionary. If they are not satisfied, they can click on the signifier 简明 to return to the *Concise* dictionary. In this way, users can switch between the two dictionaries as illustrated in Figure 2.

Figure 2:   Extract of article activated from the front page

The difference between the two types of consultation described above calls for a terminological clarification. The dictionary accessed from the front page by entering a word in the search box functions fundamentally like any other digital dictionary placed in a separate dictionary app or online portal. The consultation of these dictionaries also requires that their users enter a word or word string in a search box. The only difference is that learners can access the dictionary directly from the learning app and get lexicographical assistance without leaving the app, thus skipping a time-consuming external consultation process every time they experience a lexicographically relevant information need. Dictionaries with these characteristics can be defined as *embedded dictionaries*. By contrast, dictionaries that are activated by clicking on a word in the course text can be defined as *integrated dictionaries*. They differ from the former in that only the words occurring in the text can be consulted. Although the two types of dictionaries used in the *Kaiyan OpenLanguage* app upload their data from the same lexicographical database, they are nonetheless different, not only in the way they work, but also in the perspectives of achieving a more personalized user service. Embedded dictionaries cannot "know" what information a user is looking for in a concrete consultation, whereas integrated dictionaries, if well designed, will be context-aware and, thus, "know" the concrete sense of a word relevant to the user. This context-awareness seems to be the

most urgent lexicographical challenge to the *Kaiyan OpenLanguage* and other similar learning apps.

As already discussed above, integrated dictionaries have some obvious advantages like personalization, efficiency, and time-saving. But the advantages largely depend on the quality of the lexicographical database and the design of the user interfaces; that is, the dictionary as presented to its target users. The two aspects, the database and the interfaces, are interwoven and must be of a high standard. If the quality of the database is low or the interfaces badly designed, users may find the consultation process frustrating, time-consuming, and even unsuccessful. In the following section, we will therefore consider both aspects when we analyze how the two dictionaries are integrated into the learning app.

4.   Design features and performance

This section will focus on the dictionary integrated into the *Kaiyan OpenLanguage* app; that is, the dictionary users can consult directly from the course text whenever they have an information need in terms of understanding. It will look at the design features of the user interfaces and discuss some of the problems observed. These problems concern lemma selection (single words, multi-word units, and compound words), treatment of polysemous words, grammatical categories, translation equivalents, data overload, and position of the pop-up window.[2]

4.1   Inconsistent presentation

The treatment of the words occurring in the learning courses is inconsistent. This problem comprises their selection and treatment as lemmata, grammatical categorization, and presentation in the pop-up window.

4.1.1   Selection and treatment of single-word lemmata

The selection of lemmata and their treatment in the incorporated dictionaries seem, at least to some extent, to be arbitrary. Some words appearing in the lessons are not presented as lemmata in the pop-up window but only in the embedded dictionary on the front page. Other words do not perform as lemmata in any of the two options, although they occur in the course texts. Examples of this are *Covid* and *buzzwords*, which do not seem to be stored at all in the lexicographical database that feeds the learning app. It goes without saying that such omissions are not helpful to the users' understanding and learning of new words.

An example of a word that can be consulted on the front page but not directly from the course text is *geeks*. It appears in the learning course in a con-

text where one of the teachers explains the word *esoteric* as "not very accessible because it is mostly made by geeks for geeks". However, when users click on *geeks*, they will get a message informing them that there is "no explanation for the moment" (☹ 暂无解释).

If they instead try to consult *geeks* by entering this word in the search box on the front page, a window will be displayed with an extract of an article with the lemma *Geeks* with capital letters (see Figure 3). *Geeks* and *geeks* are two different words. They have different spelling, grammar, and meaning. *Geeks* is "a brain game on the internet; freak, a boring person", whereas *geeks* is the plural of *geek*, which the dictionary defines as "someone who does a clumsy funny show, someone abnormal; savage; someone smart, excellent at studying, but poor in dealing with social networking" (see Figure 4). Consequently, the two words are selected as two different lemmata and presented in separate articles.

The problem is clearly not related to the lexicographical database, but to the underlying programming that does not allow the relevant items contained in the database to be displayed in the user interface. The design of the latter is at fault. An appropriate design would refer the users directly to the article *geek* when they click on *geeks* in the text or enter the word in the search box on the front page. And it would also provide a small explanatory note informing them that *geeks* is plural of *geek*.
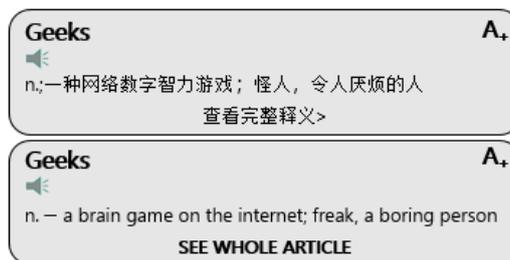
Figure 3:   The article *Geeks* accessed from front page in Kaiyan app[3]

Figure 4:   The article *geek* accessed from front page in Kaiyan app

4.1.2 Selection and treatment of multiword lexical units

In his classical *Manual of Lexicography*, Zgusta (1971: 144) wrote the following about multiword units:

> Multiword lexical units are very frequent. They carry lexical meanings as wholes (units); in this respect, they function in the sentence in the same way as those lexical units which consists of one word only.

The dictionaries used in the *Kaiyan OpenLanguage* app do not treat these frequent lexical units in a consistent and user-friendly way. Some of the units appear as lemmata in the pop-up window with equivalents and other lexicographical items when users click on them in the course text. Others cannot be looked up as a whole. For instance, when users click on *sick leave* in the text, they will not get any definition or other relevant data concerning this multiword unit as a whole but only the words *sick* and *leave* separately. However, if the users instead enter the whole unit in the search box on the front page, they will immediately get access to an article that treats *sick leave* as a normal lemma. Hence, the lexicographical database that feeds the learning app does include *sick leave* as a multiword lemma. The problem is not the database but, as above, the underlying programming that does not allow the users to get immediate lexicographical assistance in the user interface by clicking on multiword units appearing in the text.

A similar example is *dry wit*. Learners who want to know the meaning of *dry wit* can only activate *dry* or *wit* but not *dry wit* as a lexical unit. However, the combined meanings of *dry* and *wit* do not explain the meaning of *dry wit*. The users may therefore find the consultation useless and time-wasting. Other multiword units like *get the game on lock*, *pan seared*, *go ahead*, and *right now* have the same problems.

Some compound words, like *hit-and-miss* and *start-up*, cannot be consulted as a whole either. Only articles treating one of the single words will be displayed when users click on them in the text. A click on the compound *start-up* will, for instance, activate an article with the lemma *start*. However, other compounds like *self-deprecating* and *cord-cutting* can be activated and displayed as a whole (see Figure 5). This shows an inconsistent treatment of compound words in the *Kaiyan OpenLanguage* app.
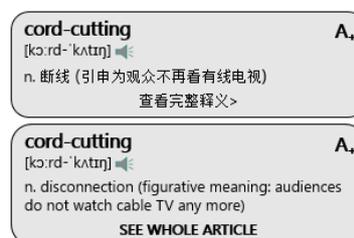
Figure 5:   Pop-up window with *cord-cutting* in Kaiyan app

4.1.3   Confusion of grammatical categories

As mentioned, some multiword units do appear as lemmata in the Kaiyan learning app now and then. In such cases, the pop-up window may contain relevant lexicographical items like pronunciation, part of speech, senses, and equivalents. But sometimes part of speech is wrongly indicated, and the equivalents that help the Chinese learners to understand the meaning do not belong to the same grammatical category as the lemma. The expression *put to good use*, for instance, is presented as a noun, although it is a verbal phrase (see Figure 6).

Figure 6:   Pop-up window with *put to good use* in the Kaiyan app

The confusion of grammatical categories may explain why the first of the two equivalents provided is the compound noun 有效使用 (efficient use), whereas the second is the much more suitable verbal phrase 用在该用的地方 (use where things should be used). This may hamper the learning process as well as the understanding of the lemma *put to good use*. In this case, the problem seems to be a low database quality originating from a poor compilation or storage of the lexicographical data that are presented in the pop-up window on demand.

4.1.4   Inconsistent presentation of senses

When a word occurs in one of the course texts, it always appears in a specific context with a specific meaning. Learners who do not understand the word in this context and activate the default pop-up window will need a definition of its specific meaning to satisfy their needs. However, the relevant senses of some words occurring in the lessons cannot be found in the *Concise Dictionary*, but only in the *Oxford English–Chinese Dictionary*. One such example is the senses addressed to the word *chunk*. When learners click on this word in the course text, an article with the senses "big block" and "stubby person or thing" will pop up (see Figure 7).
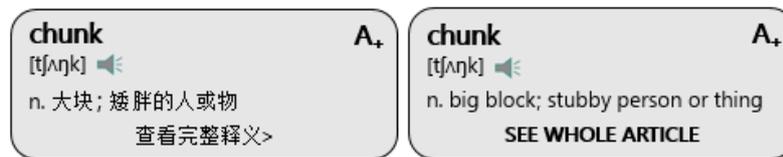
Figure 7:   Pop-up window with *chunk* in Kaiyan app

However, the senses explained in Figure 7 are not relevant in the following context quoted from the course:

> Yeah. So this is a chunk. It's a collocation. They always go together ...

If learners are aware of the problem and want a suitable definition of *chunk* as it appears in the above context, they can try to access the whole article by clicking on the signifier 查看完整释义. However, they will once again be disappointed. The article displayed is identical to the former one in terms of the senses explained in Chinese (see Figure 8). The only difference is that the whole article also offers two definitions of the noun *chunk* written in English ("a compact mass" and "a substantial amount") and one of the verb *chunk* ("put together indiscriminately"), none of them helpful in the concrete case. (Besides, definitions written in English may be difficult to understand for learners with a low English proficiency level).
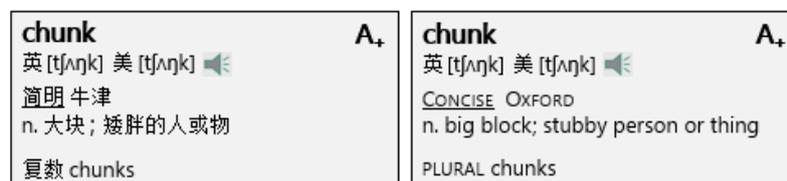
Figure 8:   Extract of article *chunk* accessed by clicking on SEE WHOLE ARTICLE

Users who do not give up can have a third try and consult the *Oxford English–Chinese Dictionary* by clicking on 牛津 (Oxford). However, as this dictionary lists *chunk* as both a noun and a verb, they are first directed to an in-between page where they have to click on one of the two word classes to proceed. A new page will finally open with an article that contains various senses of the noun *chunk*. One of these is the right one in the specific context. It defines the linguistic term *chunk* as it appears in the lesson as "a phrase or group of words which can be learnt as a unit by sb who is learning a language" and provides two examples to support this definition (see Figure 9).

chunk   **A₊**

英[tʃʌŋk] 🔊

1) a thick solid piece that has been cut or broken off sth 厚块；厚片；大块

   • A chunk of cheese/masonry
     一块厚厚的奶酪/砖石

2) (*informal*) a fairly large amount of sth 相当大的量

   • I've already written a fair chunk of the article.
     我已写出文章的大部分。

3) (*linguistics*语言) a phrase or group of words which can be learnt as a unit
   by sb who is learning a language. Examples of chunks are 'Can I have the
   bill, please?' and 'Pleased to meet you'. 话语组成部分；组块

Figure 9:   The three senses of the noun *chunk* in Oxford

A similar example is the verb *furnish*. It appears in the lesson with the meaning
"to put furniture in a house, room, etc." But the only definition in the default
pop-up window is "to offer; to supply; to equip with". Hence, the definition
does not inform the learner *with what* a house or room is equipped (furniture).
Only users who persist and consult the complete article in either the Concise
dictionary (only with a definition in English) or the Oxford dictionary (with a
Chinese equivalent) will get the required response to their information need.

The detected problems show that it makes sense to use the Oxford dic-
tionary as a safety net. The problems originate in the Concise database that
does not contain the pertinent sense of *chunk*. Four clicks to get a proper
answer may be too much for many learners who just want a quick explanation
that allows them to continue reading without interruptions. In this respect, the
interface of the default pop-up window is also to blame. It should never invite
users to consult the whole article when there is nothing relevant to add. In-
stead, they should be referred directly to the Oxford dictionary, thus skipping a
useless step that may stress them and even result in an abortive consultation.

4.2   Between data overload and data underload

Fuertes-Olivera and Tarp (2014: 64) assert that the ideal online dictionary
should contain "as much data as possible in terms of all possible consultations",
whereas its individual articles "should include as little data as possible" to
avoid information overload. From this perspective, a personalized service
requires that the dictionary only presents the exact "types and amount of data",
which the individual user needs in each consultation, "neither more nor less".
These reflections are especially relevant for dictionaries integrated into learning
apps. This type of dictionaries should be designed to offer assistance each time
the learner asks for it. But at the same time, they should avoid disturbing the
learning process with too much data. It is undoubtedly a delicate balance.

We have previously seen various examples of data underload. The default pop-up window that is supposed to provide first aid to users with comprehension problems does not address some lexical units occurring in the course texts. However, there are also problems with some polysemous words. They pose a special challenge because one or several of their relevant senses are sometimes missing. The word *buds*, for instance, appears in the lesson *The explosion of podcasts* in the following context:

> Yes, and of course, podcasts aren't the only form of media that gives you that feeling. I mean a book could also give you that feeling. But there is something about a podcast, you know, you have the buds in your ears, and it's going right into your head.

When users click on *buds* in this context, an article with three senses will visualize in the pop-up window (see Figure 10). None of them reflect the specific meaning in the context (*earbuds* or *ear headphones*). Hence, they do not help learners who do not understand *buds* in the concrete context. The relevant sense is missing. It is a case of data underload.
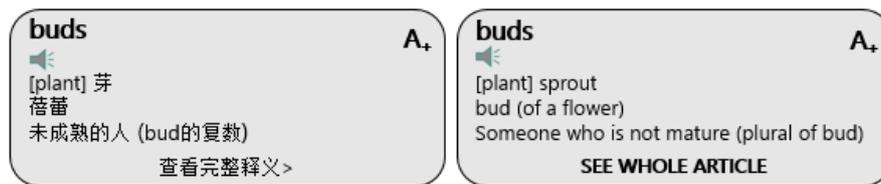
Figure 10:   Pop-up window with *buds* in Kaiyan app

Polysemous words do not only consist of two or more senses (sometimes more than ten). They frequently also belong to different parts of speech. Traditionally, there has been a vivid discussion among lexicographers on how to structure senses in dictionary articles. Several ordering principles — like chronology, markedness, frequency, and logic — have been suggested; see Lew (2013: 291-294). The problem in print dictionaries is that the structure is there for good. Digital dictionaries, at least in theory, are more flexible. However, it transpires that the dictionary integrated into the learning app has not found a technological solution to this challenge. The order of both word classes and senses seems to be unrelated to their concrete meaning in the course text. Consequently, the relevant sense and part of speech are not necessarily listed first when users click on a concrete word. The pop-up window with *deal* is a good example (see Figure 11). It has been activated by clicking on the noun *deal* in the sentence: "Can you give me a better deal?" The window contains 20 senses (or equivalents) distributed over one verb and two nouns (one of them a proper noun). This structure, together with a certain degree of data overload, will probably put users in a difficult situation when the required meaning is not the

first one listed. They will have to struggle to find their way among the many possible solutions.

Figure 11:   Pop-up window with *deal* in Kaiyan app

There are currently excellent tagging programs that can detect the parts of speech of words in a text. But they are far from perfect. Maybe five or ten percent of all words will be assigned to the wrong part of speech. It therefore makes perfect sense that the *Kaiyan OpenLanguage* app presents different parts of speech together in the pop-up window. However, this does not exclude the use of existing technology to give priority to the senses with the most likely part of speech and place them first in each case. This would, to some extent, alleviate the burden on the users, although it would not remove the general problem of data overload.
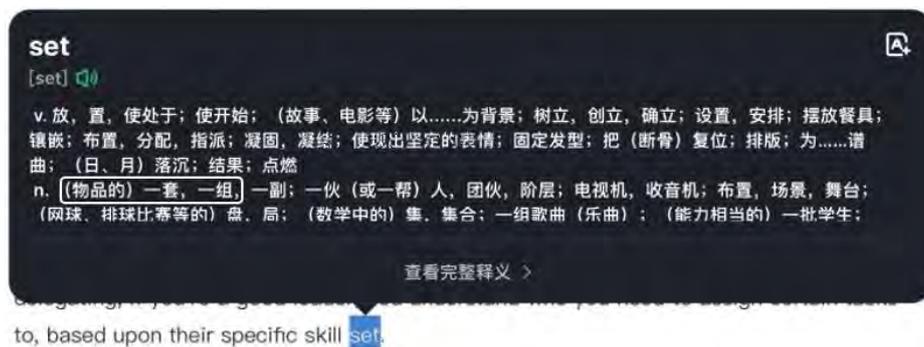
Figure 12:   Pop-up window with *set* in Kaiyan app

Many common words belong to more parts of speech and have more senses and equivalents than *deal*. One of these is *set*. Figure 12 shows the default pop-up window with this word. The white frame marks the location where the user can find its specific meaning in the given context: (物品的) 一套, 一组, which means a "suite or series, group (of things)". This meaning item is not easy to find at first glance. It will probably take several seconds to detect, evaluate, and choose the right one. The large amount of irrelevant data obstructs the information search process.
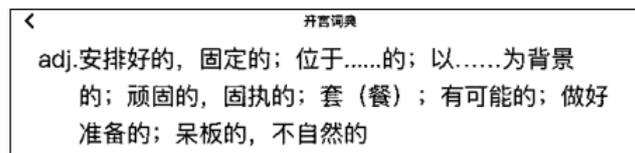
Figure 13:   The adjective *set* displayed on a new page with the whole article

So far so bad. The learners' odyssey is not over yet. Apart from being a noun and a verb, *set* can also be an adjective with several senses and equivalents. These items do not turn up in the default pop-up window shown in Figure 12. Hence, if *set* appears in the course text as an adjective, learners will have to waste even more time to find and assimilate its concrete meaning. The only way forward is to click on the signifier 查看完整释义 to access the whole article where the adjective is explained (see Figure 13). But before reaching this goal, there are plenty of possibilities that some learners will leave the consultation without any solution to their problem. They will waste their time and sacrifice their energy on the altar of data overload. A different design of the user interface is required.

4.3   Miscellaneous

Besides the challenges discussed in the previous sections, the default Concise Dictionary also presents some inadequate definitions and old-fashioned features. Most of these problems can be solved with good lexicographical craftsmanship. In this respect, the Concise Dictionary still has a long way to go to catch up with its Oxford twin in terms of quality.

Some words are explained in a way that may confuse even native Chinese speakers. For instance, one of the equivalents used to explain *topping* is 浇头, a dialect word that may be unknown to many Chinese learners. When they see this equivalent, they will have to consult a Chinese monolingual dictionary to confirm its meaning. Another example is *lanyard*. One of the equivalents provided is 系索, which is also unfamiliar to many native Chinese speakers. Even monolingual dictionaries do not include 系索 as a lemma. Learners, who look

for it in either the dictionaries used in the learning app or an external monolingual dictionary, will find nothing useful.

Moreover, some translational equivalents are inaccurate and even wrong. *Issued*, for example, is labeled as a verb, but the equivalent is a noun ("someone who issues, publishes things"). The equivalent of *gumbo* is 秋葵浓汤, which means a thick soup with okra. However, *gumbo* consists of more ingredients than just okra. The first definition of this word in the online *Longman Dictionary of Contemporary English* is "a thick soup made with meat, fish, and okra".[4]

The Chinese definition of some words has a positive connotation, although they appear in the context with a negative connotation. For instance, the first equivalent of *perpetuate* is 使不朽, which means "keep as eternal glory". This Chinese word has a positive connotation and does not seem to be an appropriate equivalent of *perpetuate* in the concrete context: "Yeah, as long as it's the truth. What I don't do is perpetuate a culturally-endorsed lie." *Perpetuate* is here used with an unfavorable meaning. Other words like *goblin* and *silverware* have similar problems that may confuse the learners.

Another problem observed in both the Concise and the Oxford dictionary is the use of abbreviations like n, v, vi, vt, adj, etc. This practice had its justification in the space constraints of paper dictionaries. But it is no longer a must in the media convergence age where these constraints have disappeared. In fact, users may sometimes find it difficult to assimilate the meaning of such abbreviations. They are an unnecessary burden on the shoulder of learners who have to confront more relevant challenges when studying a second language.

A final problem is the phonetic transcription of pronunciation, which is a leftover from the print dictionary. Traditionally, many dictionary users have struggled to convert this transcription into a more or less correct pronunciation. It is completely superfluous as a default data in a multimedia learning app like *Kaiyan OpenLanguage* where users can listen to the correct pronunciation by clicking on the speaker icon.

4.4   Position and size of the default pop-up window

The default pop-up window is situated immediately above or below the word that users have clicked on. This position is sometimes problematic as it may hide the relevant context. The users' eyes often move between the window and the context while trying to pick up the right meaning during the consultation. Figure 14 illustrates the problem. The screenshot is from a smartphone and shows how the pop-up window covers the context in which *perpetuate* occurs. This position makes it difficult for the users to grasp the negative connotation discussed above.

Figure 14:   Pop-up window with *perpetuate* in Kaiyan app

The only option for the users is to close the window to see the context and then reactivate the window to know more about the word. This switching between the pop-up window and the course text may be necessary several times until they find the right answer. It will prolong the consultation process and proba-bly stress the learners because it takes focus away from the lesson. Many of them may be tempted to stop the consultation right away. An alternative solu-tion is therefore required. From a technological point of view, there seem to be two options that go hand in hand. The first is to place the pop-up window else-where on the screen or, at the least, make it movable so the users can place it where it is most convenient. The second is to reduce its size and make it less dominant. The latter is the most complicated. It requires a combined techno-logical and lexicographical effort that aims to reduce the number of items in the window to an absolute minimum, as recommended by Fuertes-Olivera and Tarp (2014). We will discuss this challenge in Section 6.

5.   Negative effects on consultation and learning

In her study of the information search process, Kuhlthau (1991: 367) observed a whole spectrum of feelings in different phases of the process: Affective symptoms of uncertainty, confusion, frustration, and anxiety prevail when users encounter challenges and difficulties. A sense of personal inadequacy may even prompt some of them "to abandon the search altogether". These symptoms are replaced by feelings of certainty, relief, and confidence when users make progress. The study focused on information search in libraries. But similar feelings may also exist among users of learning apps with integrated dictionaries.

Kuhlthau's findings somehow converge with Krashen's Affective Filter hypothesis. According to the author of this hypothesis, various affective vari-ables "relate to success in second language acquisition" (Krashen 1982: 31). These variables include motivation, self-confidence, and anxiety. Learners gen-erally do better in second language acquisition when they are highly moti-vated, self-confident, and not anxious. By contrast, if the teaching material is

too complex or too difficult for them to assimilate, they may feel anxious and lose their self-confidence. This state of mind would certainly have negative effects on the learning process.

In the previous section, we detected several problems related to the dictionaries used in the *Kaiyan OpenLanguage* app, especially the *Concise Dictionary* that functions as the default dictionary in the pop-up window. We have seen lexical units that are not treated in the pop-up window, and sometimes not even in the dictionaries when consulted from the front page. We have seen polysemous words where some senses are missing in the default dictionary or only available after accessing the whole article. We have seen words with inaccurate and even wrong definitions. We have seen examples of data overload with senses and equivalents that are irrelevant in the concrete context. We have seen users who have to click three or four times to get an answer or no answer at all. We have seen how the position of the pop-up window that is supposed to help users sometimes has the opposite effect and make it more difficult to pick up the meaning of a word.[5]

All these problems show that the dictionaries not only assist the learners with lexicographical data. They also create all sorts of obstacles. As a result, learners will have to invest more time and energy in searching for the required information. Even so, their efforts will often be in vain. The time wasted in searching will interrupt the reading of the course text. The many fruitless look-ups will impede their understanding of the context and the learning of new words. The unpleasant user experience may cool the learners' interest in dictionaries and make them stop using them. They will become anxious and less motivated. Their English learning process will not go as smoothly as they and the app designers expected. They may even lose confidence in themselves and their ability to learn English. At worst, they may eventually stop using the learning app and give up learning English altogether.

Nobody wants this to happen, neither the designers nor the learners. Something has to be done. Dictionary consultation should be a support to language learning, not an obstacle as described above. The next section will come up with some suggestions to improve the app from a lexicographical perspective.

6.   Suggestions

The central function of the integrated dictionary is to help users understand the course texts. In this respect, learners who have problems need a quick lexicographical response to prevent them from losing focus and reading flow. The design of the pop-up window must therefore be as user-friendly and functional as possible. There is, however, also another relevant function that has to do with the complex relationship between incidental and intentional learning. Incidental learning prevails when learners gradually pick up new words, senses, and grammatical structures during reading. The pop-up window sup-

ports this part of the learning process. But learners are learners with a curious and exploring mind. When they have digested the lexicographical data displayed in the window, many of them will, from time to time, and for one reason or another, tend to go into detail about words, meaning, grammar, and cultural connotations. They will then pass from incidental to intentional learning. It thus makes sense that the learning app allows them to access the whole article either from the default pop-up window or the front page. It is a matter of course that this part of the lexicographical support should also be high quality. It represents the second layer of assistance to the users, without which the following suggestions for improvement of the pop-up window would be less attractive.

6.1   The ideal pop-up window

In the Introduction, we referred to the Chinese *Xun Gu* tradition and a similar one in Europe. This millennial experience has inspired us to design what we consider the ideal pop-up window. Figure 15 depicts how we imagine this window. It represents the lexicographical response to users who click on *set* in the context "based upon their specific skill set", which we also used in Figure 12.

Figure 15:   Content of ideal pop-up window with the noun *set*

The difference between the proposed window and the one currently used in the *Kaiyan OpenLanguage* app becomes crystal-clear if we compare Figures 15 and 12. The proposal not only contains an absolute minimum of items to meet the users' needs in the concrete context. It also breaks with well-established features of the traditional dictionary article.

The main idea is that the pop-up window should only include items that can be justified by the immediate user needs. Thus, it merely consists of a speaker icon, a meaning discriminator followed by two equivalents, and a signifier (>). The central item is the definition (or equivalents) that directly assists understanding of the course text. Other senses of the noun *set* are considered to be irrelevant in the concrete context and therefore excluded. The window also includes a speaker icon to service learners who, as recommended by language didactics, read aloud and may need to listen to some of the words to pronounce them right. Finally, it provides a widely used signifier that affords access to the whole article and is well-known to most netizens. In addition to these three compulsory items, we also envisage a note that relates a grammatical word to

its canonical form (for instance, *thought* and *think*). In some cases, this note may be sufficient for the users to understand a word appearing in the text.

In the proposed pop-up window, the lemma has disappeared. The inspiration from the *Xun Gu* tradition is crucial here. The lemma is simply superfluous. The old scribes did not need it, and neither do the digital-era scribes. When discussing a similar pop-up window in a writing assistant, Tarp (2019: 240) observes that the lemma "seems to be completely redundant as the user perfectly well knows from which word the article has been accessed." This way of excluding irrelevant items prevents data overload with negative consequences as anxiety, frustration, and abortive consultation. The design and layout shown in Figure 15 are provisional. The final design should be human-centered and optimized by test-driven development (TDD), as recommended by Norman (2013) and adapted to lexicography by Tarp and Gouws (2020). This type of optimization concerns both the presentation of the respective items in the pop-up window and the appearance, color, shape, and default position of the window itself.

6.2   How to achieve the ideal pop-up window

The method to be applied is data filtering, a technique developed within information science; see Bothma (2011). It implies that the user interfaces upload the lexicographical data from the database and offer them to users in carefully metered doses to meet their specific needs in different situations. With this technique, a single database can feed various types of dictionaries and digital products. If well-designed, it can provide relevant data to both the integrated and the embedded dictionaries in the learning app. However, even cutting-edge filtering techniques assisted by artificial intelligence cannot by themselves guarantee the reduction of relevant senses to the minimum shown in Figure 15. The minimization is also attributable to a unique characteristic of the course texts, which differ in one important aspect from those that can be found on the Internet or in e-readers like Kindle (see Bothma and Prinsloo 2013). This aspect is the limited and controlled number of words occurring in the course texts. It allows a different lexicographical approach to text reception.

Artificial intelligence contributes more and more to both language learning and lexicography. Among other things, it can detect text structures and patterns that escape the human eye, but it is nonetheless stupid as it does not understand either single words or texts as a whole. It is therefore necessary to incorporate human-assisted intelligence so that the dictionary will "know" the specific meaning of lexical units in the concrete context where they occur. How can that be done?

It requires a combination of programming and manual work. A language expert goes through the existing course texts (and the new ones that will be added in the future). This expert should be a native speaker or another person with a high proficiency level in English, preferably an experienced second-lan-

guage teacher who knows where the shoe pinches. He or she clicks on relevant lexical units, including the extended units of meaning that require special attention (see Sinclair 2010 and Rundell 2018). After the default dictionary is displayed in the pop-up window, the expert marks the right meaning items. This information is automatically recorded by a program specially developed for this purpose. When learners subsequently click on the corresponding lexical units, the pop-up window will only present the lexicographical data marked by the expert; that is, data that help them understand lexical units in their concrete context. In this respect, the tool has become context-aware. If the users want additional information, they can click on the arrow (>) to access the complete article with its conventional features.

The suggested method is highly productive. A well-trained expert will probably be able to mark two-three words or units per minute. The solution is also economical as the described program is relatively simple and easy to write.[6] Over the long term it could be supplemented by various types of user feedback. All this will greatly improve the quality of the learning app.

6.3   Interdisciplinary collaboration

The lexicographical database used to sustain the *Kaiyan OpenLanguage* app was probably compiled before the app; that is, without considering the specific requirements of this tool. It would explain many of the deficiencies observed. Fuertes-Olivera and Tarp (2020: 268) discuss a similar problem related to a writing assistant. They express serious doubts about the capacity of existing databases to feed innovative digital tools like writing assistants and conclude that "it seems necessary to restart almost from scratch". We can recognize many of their observations and reservations. We nevertheless believe that it is possible to improve the database that feeds the *Kaiyan OpenLanguage* app as long as there is a close collaboration between the different stakeholders.

The analysis of the integrated dictionary in Section 4 revealed several cases where single words, multiword units, and specific senses were missing in the database, although they occurred in the course texts. When the language expert who scrutinizes the texts discovers these omissions, they should be marked and sent to the compilers of the dictionary for immediate response. Depending on the work burden, the new items could be ready and activated within 24 hours. This requires close interdisciplinary collaboration between the app designers and the lexicographers responsible for the database.

The described collaboration and interaction may not be possible in terms of the Oxford dictionary. But it is probably much easier with the Concise dictionary as we consider the same company to be behind both this dictionary and the learning app. If it is not the case, agreements on close interdisciplinary collaboration should be made between companies when licensing a dictionary to be used in digital tools like the learning app. The database must necessarily be dynamic and under continuous development, improvement, and updating,

among other things because new texts are regularly added to the course. If not, the company will never get high-quality lexicographical support for its learning app.

7.   Conclusions

Digital applications to assist second-language learning are becoming increasingly popular in China and elsewhere. These apps typically incorporate one or two dictionaries to improve the service and avoid that users leave the app to look up in external dictionaries. In this contribution, we have analyzed the dictionary integrated into the *Kaiyan OpenLanguage* app, which is one of the most comprehensive app used by Chinese learners of English. The study has detected a number of inconsistencies and deficiencies such as words, multi-word units, and senses that occur in the course texts but are not treated in the pop-up window or are difficult to find because of data overload. These problems may impact negatively on learners' motivation and the learning process as such. Thus, we have suggested an alternative, context-adapted approach inspired by the work of classical Chinese and Greek scribes. The content of the proposed pop-up window has been reduced to an absolute minimum. The idea is to avoid a consultation process that interrupts the learners' focus and reading flow.

The study focuses on a particular Chinese learning app. Yet, we believe that its main conclusions and suggestions can be generalized, first and foremost to be adapted to other English learning apps for Chinese users. But it is probably also interesting for designers of apps that assist learners of English with other mother tongues as well as the learning of other languages.

Acknowledgements

Endnotes

1.   Translated from the website: https://www.OpenLanguage.com/p/about/.

2. Our observations are based on numerous consultations between October 20, 2020, and January 14, 2021. Some of the discussed examples may have changed after being consulted.

3. The pop-up window is redesigned to facilitate reading. In this and other figures, an English translation has been added for the benefit of non-Chinese speaking readers.

4. Accessed January 9, 2021.

5. One of the authors has used the app more than three months and recognizes some of the adverse feelings. The idea to the article was triggered by this experience.

6. This was confirmed by Kasper Fisker, a computer scientist at the Center of Excellence in Language Technology at Ordbogen A/S, Denmark.

References

A.   Digital dictionaries

Concise Dictionary in Kaiyan OpenLanguage.

Longman Dictionary of Contemporary English Online. https://www.ldoceonline.com.

Oxford English–Chinese Dictionary in Kaiyan OpenLanguage.

B.   Other literature

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

Bothma, T.J.D. and D.J. Prinsloo. 2013. Automated Dictionary Consultation for Text Reception: A Critical Evaluation of Lexicographic Guidance in Linked Kindle e-Dictionaries**.** *Lexicographica* 29(1): 165-198.

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

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

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

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

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.

Kang, S. 2020. Thoughts on a Chinese Learner's Dictionary of Media Convergence: A Case Study of the Contemporary Chinese Learner's Dictionary. *Journal of Ludong University* 37(2): 1-7.

Krashen, S.D. 1982. *Principles and Practice in Second Language Acquisition.* Oxford/New York: Pergamon.

Kuhlthau, C.C. 1991. Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science* 42(5): 361-371.

Lew, R. 2007. Linguistic Semantics and Lexicography: A Troubled Relationship. Fabiszak, M. (Ed.) 2007. *Language and Meaning: Cognitive and Functional Perspectives*: 217-224. Frankfurt am Main: Peter Lang.

Lew, R. 2013. Identifying, Ordering and Defining Senses. Jackson, H. (Ed.) 2013. *The Bloomsbury Companion to Lexicography*: 284-302. London: Bloomsbury.

Li, Y. and D. Wang. 2020. Several Trends in the Historical Development of Chinese Dictionaries. *Journal of Ludong University* 37(1): 1-6.

Loewen, S., D. Crowther, D.R. Isbell, K.M. Kim, J. Maloney, Z.F. Miller and H. Rawal. 2019. Mobile-Assisted Language Learning: A Duolingo Case Study. *ReCALL* 31(3): 293-311.

Ma, Q. 2019. University L2 Learners' Voices and Experience in Making Use of Dictionary Apps in Mobile Assisted Language Learning (MALL). *International Journal of Computer Assisted Language Learning and Teaching* 9(4): 18-36.

McArthur, T. 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer.* Cambridge: Cambridge University Press.

Norman, D. 2013. *The Design of Everyday Things*. New York: Basic Books.

Rosell-Aguilar, F. 2018. Autonomous Language Learning Through a Mobile Application: A User Evaluation of the *busuu* App. *Computer Assisted Language Learning* 31(8): 854-881.

Rundell, M. 2018. Searching for Extended Units of Meaning — and What to Do when You Find Them. *Lexicography — Journal of Asialex* 5(1): 5-21.

Sinclair, J.M. 2010. Defining the Definiendum. De Schryver, G.-M. (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*: 37-47. Kampala: Menha Publishers.

Stathi, E. 2006. Greek Lexicography, Classical. Brown, K. (Ed.). 2006. *Encyclopedia of Language and Linguistics. Vol. 5*: 145-146. Second Edition. Oxford: Elsevier.

Tarp, S. 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249.

Tarp, S., K. Fisker and P. Sepstrup. 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.

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.

Wurman, R.S. 1990. *Information Anxiety*. New York: Bantam.

Yong, H. and J. Peng. 2008. *Chinese Lexicography. A History from 1046 BC to AD 1911.* Oxford: Oxford University Press.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Zhang, Y. 2019a. On the Innovation of Dictionary Compilation and Publication in the Context of Media Convergence. *Research on Language Strategies* 4(6): 79-89.

Zhang, Y. 2019b. Opportunities and Challenges for Dictionary Compilation and Publication in the Age of Media Convergence. 2019/05/09. *Language and Writing*: 1-3.

Zhang, J. and Y. Jin. 2017. Current Status of Mobile Language Learning Apps in China. *Journal of Yunnan Normal University* 49(3): 22-29.

Peter Martin. *The Dictionary Wars: The American Fight over the English Language*. 2019, xii + 358 pages. ISBN 978-0-691-18891-1. Princeton: Princeton University Press. Price: $14.61 (Paperback) / $29.95 (Hardback).

English lexicography broadly encompasses dictionaries of British English, American English, and the varieties of English spoken in other countries (Cowie 2009: 1). Though the evolution of English lexicography owes much to the progression of lexicographic theory and practice in and outside Britain, the exploration of lexicography in other English-speaking countries, for example American lexicography, is conducive to a panoramic vision of English lexicography. Nonetheless, little research has been conducted in this field.

The Dictionary Wars: The American Fight over the English Language, which may be the first thorough investigation of the early period of American lexicography, delves into the history of the fierce competitions among lexicographers and the intermittent national conflicts on the way to making the most authoritative dictionary of American English in the early 19th century. It reflects 'America's progress and struggle with the English language, mediated by the country's ongoing dictionary controversies' (p. ix).

This book consists of 17 chapters in two parts, plus a conclusion, four appendices, some notes and an index. The two parts are organized in sequence from the publication of the first dictionary *A Compendious Dictionary of the English Language* (1806) by Webster to the appearance of the royal quarto edition of *An American Dictionary of the English Language* (1864). The chapters in each part progress in order of time with the story around two main characters — Noah Webster and Joseph Emerson Worcester.

Part One 'Noah Webster's Battles' depicts, in nine chapters, the cultural conflicts between Britain and America as well as the rivalry between Webster and other American lexicographers in their efforts to compile dictionaries of American English. Chapter 1 gives a brief account of the dilemma that Americans faced after political independence from Britain. On the one hand, in order to promote national identity and confidence, there was a growing consciousness that American English should be freed from British English constraints and conventions. On the other hand, most literate and cultured Americans put heavy reliance on Samuel Johnson and his dictionary. Though the practice of 'Americanisms' had suffered acrimonious attacks from Britain, an American dictionary was a must for following the rapid changes in the American language.

Chapter 2 introduces the life of Noah Webster, and the reason why Webster decided to publish his own spelling book. Webster wanted to prevent the corruption of the language from England. It was his belief that the reform in spelling could purify the American language. The publication of his speller in 1783 turned out to be a great success. Webster also contributed to federal copyright protection, resulting in the passing of the Copyright Act in 1790. As a provocative and courageous man, Webster publicly attacked, in his *Disserta-*

tions on the English Language (1789), Samuel Johnson as well as Johnson's dictionary, and maintained that the establishment of a national standard language should be an indispensable part of a complete American cultural revolution.

Chapter 3 recounts how Webster compiled his first dictionary. In 1798, Webster and his family moved to New Haven, where he was determined to compile a dictionary to fulfill his plan for education. By pointing out the errors and ignorance of British grammarians, Webster highlighted the moral role that grammar and lexicography could play in language education. According to him, vulgar words should be omitted, and dictionary definitions should be morally instructive. In 1806, Webster's first dictionary, *A Compendious Dictionary of the English Language*, came out. The innovations of the dictionary lie in its recommended spellings, its lexicon containing more than five thousand entry words not in Johnson's dictionary, and its inclusion of several American tables as well as lists for special communities. However, some inconsistencies with respect to spelling and pronunciation were found in the dictionary, which went against Webster's will of pursuing the linguistic uniformity for American English.

Chapter 4 demonstrates Webster's perseverance and concentration on lexicography. It begins with the depiction of the difficulties that Webster encountered. It took him almost ten years to sort out the etymology of the language which was particularly inadequate in preceding dictionaries. Though the German linguists had made some achievements in this field, Webster did not refer to any of them. 'Webster's work in etymology illustrates the extreme isolation and provincialism of American scholarship in the early years of the nineteenth century' (Krapp 1966: 365). Despite attacks, Webster still held the firm belief that Americanisms should be included in a dictionary. In December 1828, Webster's unabridged dictionary, *An American Dictionary of the English Language*, was published by Sherman Converse. Though Webster's spelling innovations are controversial, most of them have become the standard in American orthography. Furthermore, Webster's way of defining words distinguishes itself from other dictionaries. Even Murray, the chief editor of the *Oxford English Dictionary*, called Webster 'a born definer of words' (Murray 1900: 43).

Chapter 5 reveals how the octavo abridgement of Webster's 1828 quarto was made. With the rapid development in printing and expansion in middle-class readership, both Webster and Converse felt it necessary to abridge the 1828 quarto into a compact octavo edition which would be more suitable for average users. However, Webster himself did not want to get involved in this work. Converse decided to invite Joseph Emerson Worcester, who was famous for his scholarship in the revision of the Todd-Johnson dictionary, to undertake the labor. Under the persuasion of Converse and Goodrich, Worcester finally accepted their invitation. Worcester suggested that some modifications in Webster's quarto should be made, such as the omission of the spelling oddities, the alphabetical order of words according to conventional spelling, and the shortening of definitions.

Chapter 6 briefly describes how Goodrich, his son-in-law, took over the

copyright of the octavo abridgement. As expected, the publication of the one-volume octavo in 1829 was a success. However, the revisions in spelling, pronunciation and definitions made Webster extremely angry because he thought all these would damage his image as America's premier authority on the English language. As a result, Webster decided to sell the copyright of the octavo to Goodrich. After receiving both the copyright and the plates in 1833, Goodrich found a new publisher — the Whites — who would retain exclusive publishing rights for the ensuing twenty-four years.

Chapter 7 shows the contest between Webster and Lyman Cobb with their spelling books. As a young schoolteacher, Cobb published his own speller adopting Walker's system of pronunciation in 1821. Cobb anonymously launched several rounds of attacks on Webster by pointing out that Webster failed to achieve consistency in spelling and pronunciation between his spelling books and dictionaries. In response, Webster reiterated that his goal was to purify the American language and unify the United States. The so-called 'spelling wars' between them for the first time made the Americans aware of the irreconcilable spelling differences between America and Britain.

In Chapter 8, 'The "Common Thief"', the author vividly depicts the fierce competition in dictionary compilation between Webster and Worcester. Worcester's *A Comprehensive Pronouncing and Explanatory Dictionary of the English Language* (1830), which was widely accepted, posed a formidable challenge to Webster. The merits of Worcester's dictionary lie in its utilitarian values with a lower price, concise definitions, and emphasis on pronunciation with various citations from 18th century orthoepists. However, Webster claimed that such treatment of pronunciation was a retrogression — it opposed his mission for a uniform language. By listing several areas where he thought Worcester had committed fatal mistakes, Webster brought the dispute to the general public. To make the dispute more furious, Webster published an article in a newspaper charging Worcester with plagiarism from his dictionary. This article 'marks the first time the quarrels between the two men had broken out into a public forum' (p. 131), which can be seen as the start of the dictionary wars. In the next few years, several rounds of attacks proceeded between the two lexicographers. According to the author, the quarrel is connected, to some extent, with their personalities. While Webster is a radical language reformer, who was inclined to maintain his exclusive status as a leading lexicographer in America, Worcester was a modest conservative, who strongly believed that the linguistic custom or prevalent usage should never be violated. However, the dictionary wars helped to raise the American consciousness for the significant role that lexicographers were beginning to play in American society.

Chapter 9 briefly introduces Webster's last dictionary and some comments on his work. Because of his worry of being overcome in the competition, Webster decided to publish a new dictionary by making all the revisions and corrections needed in his first quarto dictionary. While continuing his anti-British theme, he still refused to retreat from his mission to purify the American lan-

guage and promote linguistic reform. With the help of Goodrich, Webster's second dictionary *An American Dictionary of the English Language* or the royal octavo was published in 1841. However, except for the additional words, there is no big difference between this royal quarto and the first quarto. Webster died in 1843. His patriotism to America, devotion to lexicography, and innovations in linguistic reform have made him widely recognized and won him much credit.

Part Two 'The Merriams at War' describes, in eight chapters, the language wars for supremacy after Webster's death. In Chapter 10, 'Taking Webster out of Webster', the author makes a detailed analysis of the strife among Webster's family members and the new move taken by the Merriam brothers. Misunderstandings and resentments existed among Webster's sons-in-law in respect of the executorship of Webster's estate. As the dominant executor, William Ellsworth decided to find a purchaser for Webster's unsold sheets with the expectation of challenging Worcester's new dictionary. It was the brothers Charles and George Merriam who finally bought those sheets. Gifted with a printing business and being highly aware of the profit that the name 'Webster' would bring about, the Merriams wanted to publish a new edition of the royal octavo immediately. Goodrich was once again recommended as the editor, and he demanded that the new Merriam edition should be coordinated with revisions of the other Webster dictionaries. Finally, an agreement was reached that the revisions of the Merriam edition and the Goodrich-Worcester octavo should be kept in step with each other with regard to spelling, pronunciation, definitions, and etymology. A massive project as it sounds, 'Goodrich's plan made possible the thorough housecleaning required throughout Webster's lexicography' (p. 177).

Chapter 11 illustrates the strategies that the Merriams took in competing against Worcester and his dictionary. In 1846, Worcester published his *Universal and Critical Dictionary of the English Language*, the most outstanding feature of which is its complete inclusion of the vocabulary in use. This dictionary was produced in a context where 'words were no longer the concern chiefly of the learned and social elite' (p. 182). Having realized the threat from Worcester, the Merriams began to look for evidence of Worcester's alleged thefts. They asked William Webster to search the words that Worcester borrowed from Webster's dictionaries, and called on Noah Porter to launch an attack against Worcester. Their goal was to disparage Worcester and advertise their forthcoming dictionary. The one-volume Goodrich-Merriam quarto edition of the *American Dictionary* was published in 1847. This new edition stands for a significant turning-point in the history of Webster dictionaries: the innovations proposed by Webster were excluded; and the function of an American dictionary was redefined: 'to help people from all walks of life understand how to speak and write the language' (p. 193).

Chapter 12 discloses in detail the publishers' intrigues in the dictionary market. The story begins with Worcester's discovery of his name appearing with that of Webster in his 1846 dictionary published in London in 1851. In fact,

it was the London publisher Henry G. Bohn's deliberate mistake by changing the title to better promote the dictionary. Then the Merriams took advantage of this fraud and spearheaded attacks on Worcester again. Despite Worcester's struggle urging Converse and Goodrich to clarify the matter, he eventually failed. On the one hand, this fraud reflects that 'publishers tend to make exaggerated, unsubstantiated, and even false claims about their dictionaries either in the history of lexicography or in modern business practice' (Landau 2009: 197). On the other hand, it is through this kind of conflict that the publishers could achieve their goal of attracting the attention of the American public to the dictionaries.

In Chapter 13, 'Converse's Complaint', the author briefly covers Converse's involvement in the dictionary wars between the Merriams and Worcester. As the publisher of the Goodrich-Worcester octavo dictionary, Converse accepted Worcester's invitation to set the record straight. However, he ended by failing to gain even Goodrich's support.

Chapter 14 continues the account of the rivalry between the Merriams and Worcester. In 1854, Worcester's new publisher Swan, who was a mason turned schoolteacher and a school textbook publisher, initiated a comprehensive rebuttal of the Merriams. According to Swan, the fatal flaw of Webster's dictionary was his spelling innovations, which would perplex schoolchildren. The involvement of schoolchildren into this debate raised the tone of the controversy because it was the age of child education. In spite of the great improvements that the Goodrich-Merriam quarto edition had made, the Merriams were aware of their weaknesses, particularly Webster's orthography. Therefore, the Merriams had to seek the endorsement from public figures. However, most leading authors and distinguished men of letters favored Worcester.

Chapter 15 underscores the topic of 'standard language' in the dictionary wars. In a reply to the Merriams' attack, Swan cited William Cullen Bryant's criticism on Webster's alleged nationalism, and argued that 'the English language was an apolitical heritage and should be adapted to the realities of American life' (p. 236). Then the debate began centering on the 'American standard'. In 1855, Worcester's octavo, *A Pronouncing, Explanatory and Synonymous Dictionary of the English Language*, was published, in which he put forward his own understanding of 'standard', namely that the standard lies in the present usage of a literate and well-bred society.

Chapter 16 shows another influential dictionary made by Worcester. During their preparations for the new quarto, Worcester's publishers had inadvertently spilled the beans. Their innovations of including woodcut illustrations had been borrowed by the Merriams, who published the Pictorial Edition of *An American Dictionary of the English Language* in 1859. Whatever strategies the Merriams took to win over the market, Worcester's magnum opus, the two-volume, unabridged quarto, turned out to be a triumph. In the preface to the dictionary, he stated his principal tenets in lexicon compilation. Worcester died in 1865 and his battles finally came to an end.

In Chapter 17, 'The Merriams Triumphant: "Worcester! Worcester! All Change for Webster!"', the author deals with the realization of the royal quarto edition published by the Merriams. Though faced with the economic depression caused by the Civil War, the Merriams still decided to forge ahead. They armed their working team with thirty professors, scientists, and general assistants. With their joint efforts, the royal quarto edition of *An American Dictionary of the English Language* came out in 1864. With a comprehensive revision in etymology, spelling, pronunciation, and definition, this dictionary proved to be a great success. Also, it marks the ending of the American dictionary wars during the first half of the 19th century.

In the concluding section, the author expresses his own views on the dictionary wars: neither Webster nor Worcester is the real winner. However, their influence on American language and culture will last forever. The real winners, in the author's view, are the Merriams, the newspapers, and the American people. As for the Merriam brothers, they had taken the lion's share in the dictionary market and earned themselves a fortune. As for the newspapers, they had become center stage during the dictionary wars. As for the American people, their knowledge of the issues apposite to the language and the dictionaries made them the real beneficiaries.

The back matter contains four appendices. Appendix A, 'The "Webster" Brand', briefly introduces another dictionary conflict related to the brand name 'Webster'. By 1909, the name 'Webster' had become a genericized trademark for any American dictionary and a symbol of quality for American readers. Appendix B, 'Four Centuries of Selected Dictionaries of the English Language', presents a list of selected general-purpose or monolingual dictionaries of the English language from the 17th to the 20th century. The dictionaries are arranged chronologically, and each entry comprises the publication date, the editor(s), and the dictionary title or a description of the dictionary. Appendix C, 'Publishing Terms', gives a brief explanation of publishing terms. These terms, each followed by a definition, are alphabetically listed. Appendix D, 'The Spelling Bee at Angels' quotes a poem narrated by Truthful James. It is about a spelling match, an American educational innovation. This part may be regarded as a supplement to American lexicography.

As the first publication to unveil the early period of American lexicography, *The Dictionary Wars* exhibits several merits.

Firstly, unlike other books dealing with similar topics that give a very rough and general introduction to the whole history of lexicography in a given country (e.g., Cowie 2009; Considine 2019), this book pays specific attention to a particular period of lexicography in America, which is bound to provide a more in-depth account of how American lexicography has developed and what makes it what it is today (cf. Zhang and Xu 2020). Secondly, this book is characterized by its impressive breadth of research, including the fields of education, law, religion, and publishing. All these were interconnected and had a role to play in accelerating the American cultural revolution, particularly by

promoting American dictionaries. In this way, the author spotlights the social nature of language as well as its multifaceted influence on the production of dictionaries. Thirdly, despite the large number of dramatic and complex personal stories in the book, the author manages to re-create the historical events and set the record straight through an impartial and objective description. As the author expresses it: 'In an extended, complicated, and comprehensive conflict such as this, lasting some thirty-five years, there were many victories and defeats for all the main participants along the way' (p. 288). Finally, the title of the book and of each chapter, such as 'The Dictionary Wars', 'Tea and Copyright' and 'The "Common Thief"' which suggest lively stories behind them, are eye-catching and will arouse readers' interest as well as increasing the readability of the book.

Admittedly, apart from the strengths mentioned above, there are still some possibilities for improvement of the book. To begin with, the author could have drawn more on other lexicographic resources so as to better illustrate how American lexicography influences and is influenced by lexicographic theory and practice from other countries. Except for the brief reference to the etymological discoveries by German linguists, the resources the author turns to are basically limited to British lexicography. Another minor criticism is the neglect of an introduction to lexicographic theories that guide dictionary-making. Though the focus of this book is on lexicographic practice, it is still necessary to refer to the framework on which a given dictionary is based. In addition, it would be more reader-friendly if the author had made a systematic comparison of the major dictionaries mentioned in this book, through which readers will acquire a better understanding of the development of American lexicography. Except for some brief explanations of the features of the dictionaries, the author did not make any diachronic or synchronic comparison between the leading American lexicographers and their dictionaries.

Overall, despite the minor criticisms raised above, *The Dictionary Wars* deserves recognition. It is a valuable resource for the study of English lexicography as a whole as well as for American literature and communication. The book takes readers on a linguistic journey into American dictionaries from the early 1800s to the late 1860s, into the cultural wars in which America's colleges, libraries, newspapers, religious groups, and state legislatures actively engaged, and into a business world where publishers all vied for dictionary supremacy. *The Dictionary Wars* has set a good example for an accurate and thorough assessment of the evolution of lexicography.

References

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

Cowie, A.P. 2009. Introduction. Cowie, A.P. (Ed.). 2009: 1-14.

Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography*. *Volume I. General-purpose Diction-aries.* Oxford: Oxford University Press.

Krapp, G.P. 1966. *The English Language in America*. New York: Frederick Ungar.

Landau, S.I. 2009. Major American Dictionaries. Cowie, A.P. (Ed.). 2009: 182-229.

Murray, James A.H. 1900. *The Evolution of English Lexicography.* Romanes Lecture. Oxford: Claren-don Press.

Zhang, X. and H. Xu. 2020. Review of *The Cambridge World History of Lexicography. Lexikos* 30: 629-634.

Gong Cheng
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou
China
(chenggong0128.love@163.com)

and

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

Feminine Personal Nouns in the Polish Language. Derivational and Lexicographical Issues

Agnieszka Małocha-Krupa, *Institute of Polish Philology,*
Philological Faculty, University of Wrocław, Poland
(agnieszka.malocha-krupa@uwr.edu.pl)

Abstract: The paper is dedicated to issues related to designations of women in the Polish language from the second half of the 19th century until the present time. The socio-cultural history of a group of Polish feminine personal nouns (referred to as *feminitives* or *feminatives*) which denote women's social and/or occupational status is discussed. It is argued that feminine personal nouns have been directly dependent on various ideologies: women's emancipation, socrealism and feminism. Ideologies have impacted the use of feminatives, by intensifying or limiting their use in discourse during a particular period, and the attitude of language users to ideologies has influenced the way in which feminatives are perceived. While presenting the richness of the repertoire of gender exponents in contemporary Polish, the possibility of the incorporation of feminine personal formations into dictionaries of general Polish in a scientific and objective manner is investigated. A similar idea was proposed at Wrocław University, as a result of which a group of female lexicographers compiled *Słownik nazw żeńskich polszczyzny* [Dictionary of Polish Female Nouns]. Some of its innovative lexicographical assumptions (description, not prescription, a discourse-centred method) are discussed in this article. The text corpus presented in the article enables the reader to trace the history of feminine personal nouns in Polish, i.e. their disappearance and re-appearance in the language.

Keywords: FEMINATIVUM/FEMININE PERSONAL NOUNS, WORD FORMATION, LEXICOGRAPHY, LANGUAGE CULTURE, EMANCIPATION OF WOMEN, FEMINISATION, GENDER MORPHOLOGY, GENDER AND LANGUAGE, DERIVATION, PRAGMATICS

Opsomming: Vroulike persoonsname in Pools. Afleidings- en leksikografiese kwessies. Hierdie artikel skenk aandag aan kwessies rondom die benoeming van vroue in Pools vanaf die tweede helfte van die 19de eeu tot vandag. Die sosio-kulturele geskiedenis van 'n groep Poolse vroulike persoonsname (waarna verwys word as *feminitiewe* of *feminatiewe*) wat vroue se sosiale en/of beroepsstatus aandui, word bespreek. Daar word aangevoer dat vroulike persoonsname direk afhanklik was van verskeie ideologieë: die emansipasie van vroue, Sosialistiese Realisme en feminisme. Ideologieë het die gebruik van feminatiewe deur middel van die intensivering of beperking van hul gebruik in die redevoering in 'n bepaalde periode beïnvloed, en die taalgebruikers se houding teenoor ideologieë het 'n uitwerking gehad op die manier waarop feminatiewe waargeneem word. Terwyl die rykheid van die repertoire van gendervoorbeelde in

hedendaagse Pools bespreek word, word die moontlikheid van die insluiting van vroulike per-
soonsnaamvorme in algemene Poolse woordeboeke op 'n wetenskaplike en objektiewe manier
ondersoek. 'n Soortgelyke idee is by die Wrocław Universiteit voorgestel waarna 'n groep vroulike
leksikograwe die *Słownik nazw żeńskich polszczyzny* [Woordeboek van Poolse vroulike naam-
woorde] saamgestel het. Sommige van die innoverende leksikografiese aannames (deskriptief, nie
preskriptief nie, 'n diskoersgesentreerde metode) word in hierdie artikel bespreek. Die tekskorpus
wat in die artikel voorgelê word, stel die leser in staat om die geskiedenis van vroulike persoonsname
in Pools, m.a.w. hul verdwying en herverskyning in die taal, na te spoor.

Sleutelwoorde: FEMINATIVUM/VROULIKE PERSOONSNAME, WOORDVORMING,
LEKSIKOGRAFIE, TAALKULTUUR, EMANSIPASIE VAN VROUE, FEMINISASIE, GENDER-
MORFOLOGIE, GENDER EN TAAL, AFLEIDING, PRAGMATIEK

1.   Introduction

Feminine personal nouns, their stylistic value, normative assessment, and — more broadly — ways of expressing information about the feminine gender in the Polish language have been debated on for over 120 years. The earliest discussions on the topic are known, for instance, from linguistic periodicals from the beginning of the 20th century: *Język Polski* [The Polish Language] and *Poradnik Językowy* [A Language Guide], and the most recent analyses have been presented by Karamańska and Młynarczyk (2019), Kiełkiewicz-Janowiak (2019), Szpyra-Kozłowska (2019a, b). During the last twenty years, quite a few linguistic monographies have been devoted to the complex issue of feminine personal nouns in Polish, including: Małgorzata Karwatowska, Jolanta Szpyra-Kozłowska: *Lingwistyka płci. Ona i on w języku polskim* [Gender Linguistics. She and He in the Polish Language] (2005); Marek Łaziński: *O panach i paniach. Polskie rzeczowniki tytularne i ich asymetria rodzajowo-płciowa* [On Men and Women. Polish Titular Nouns and Their Sex/Gender Asymmetry] (2006); Marta Nowosad-Bakalarczyk: *Płeć a rodzaj gramatyczny we współczesnej polszczyźnie* [Natural and Grammatical Gender in Contemporary Polish] (2009); Katarzyna Dembska: *Tendencje rozwojowe polskich i rosyjskich nazw zawodowych kobiet na tle języka czeskiego* [The Development Tendencies in Polish and Russian Female Profession Names within the Context of the Czech Language] (2012); and Agnieszka Małocha-Krupa: *Feminatywum w uwikłaniach językowo-kulturowych* [Feminitives and Their Linguistic and Cultural Entanglements] (2018). For a very long time now, feminine personal nouns have belonged to the standard repertoire of issues igniting contentious, frequently affective debate, dependent on numerous extra-linguistic parameters. That is why on 25 November 2019 the Council for the Polish Language (a consultative and advisory institution with regard to the use of the Polish language, established in 1996 by the Executive Board of the Polish Academy of Sciences (PAN)) published a communique on feminine forms of names of professions and titles, in which they state as follows:

> The right to use feminine personal nouns should rest with the speaker, and one should remember that along with recently publicised calls in the media for creating them, there exists some resistance against them. Not everyone will refer to a woman as *gościni* [a hostess] or *profesorka* [a professoress], even if she herself explicitly voices such an expectation. The Council for the Polish Language at the Executive Board of PAN recognises that Polish needs a more expanded, possibly full symmetry of masculine and feminine personal nouns in its vocabulary. (http://www.rjp.pan.pl/index.php)

In my research — to which this article refers — I apply the notion of *feminativum* both to the category of *nomina feminativa* (an open set of forms) and its individual textual representations. I define the latter as a linguistic form that is a synthetic (one-word) name of a woman, containing a morphological (suffixal or paradigmatic) indicator of femininity. My reflection, which arises from more than a dozen years' excerption of feminine material and its scientific examination, primarily focuses on designations that are synthetic names of female professions, titles, academic degrees, social statuses (sometimes I also analyse feminine possessive names, surnames of wives and daughters, as well as attributive names).

2.   How we explicate information about gender in the Polish language

From the perspective of the word-formation system of the Polish language, feminine personal nouns are derived from masculine derivational bases by taking various suffixes: the most productive *-ka* (e.g. *nauczyciel→nauczyciel-ka* [a male/female teacher], *student → student-ka* [a male/female student]), *-ini/ -yni* (*przodek→przodk'-ini* [a male/female ancestor], *wydawca→wydawcz-yni* [a male/female publisher]); as well as some less frequent ones such as: *-ica/-yca* (*anioł→aniel'-ica* [a male/female angel]), *-ina/-yna* (*hrabia→hrab'-ina* [a count/a countess]), *-anka* (*kolega→koleż-anka* [a male/female friend]), etc. Nowadays, more and more feminine personal nouns are being formed by adding the paradigmatic formative *–a* to masculine forms such as *przewodniczący, księgowy, alimenciarz, doktor, maestro, minister*. This is how the traditional uses were formed: *przewodnicząc-a* [a chairwoman], *księgow-a* [a female accountant], as well as: *alimenciar-a* [a non-working woman who lives off maintenance money], *doktor-a* [a female doctor], *maestr-a* [a female maestro], *ministr-a* [a female minister], which are to be found in discourses of pro-gender-equality circles.

The trend to designate women in a way that communicates information about their gender derivationally, which has been discussed above, is competing with the concealment of that information in generic forms such as: *doktor* [a male/female doctor or physician], *klient* [a male/female customer], *minister* [a male/female minister], *pacjent* [a male/female patient], *prezydent* [a male/female president] that are used to refer to both genders. There is still another trend — an intermediate, universally approved one, which consists in creating analytic constructions: the form of address *pani* [madam] combined with a noun which

is shared by the genders, e.g. *pani minister* [madam minister], *pani doktor* [madam doctor], *pani prezydent* [madam president]. The above nominating techniques, i.e. derived personal feminine nouns, the use of gender neutral nouns, and the use of analytic constructions are competing with each other, with the second and third models being predominant in the case of jobs, professions, functions and academic degrees regarded as prestigious. Since 1989 (a period of political and cultural transformation and a transition from a communist to a democratic system in Poland), the models have been attracting criticism from persons from pro-gender-equality circles. In line with the idea of gender-and-sexual symmetry in language, which they promote (Karwatowska and Szpyra-Kozłowska 2005), they propose that the feminine form should be derived from each masculine one, and they are using more and more synthetic feminine personal forms; often those are innovative constructions, not accepted by the customary norm, e.g. *detektywka* [a female detective], *geniuszka* [a female genius], *hydrauliczka* [a female plumber], *kibicka* [a female fan], *krótkowidzka* [a short-sighted woman], *majorka* [a female major], *osiedleńczyni* [a female settler] or *rabusia* [a female robber], and falling outside codified Polish. Apart from the existence of systemic, derivational solutions, information about gender in Polish can also be expressed inflectionally, by combining the exponent of femininity with inflected parts of speech:

— an adjective: *nową podsekretarz* [a new female undersecretary], *znakomitą minister* [an excellent female minister], *bądź wrażliwą* [be sensitive (spoken to a woman)];
— an inflected participle: *wybieraną przedstawicielka* [an elected female representative], *interesującą kobieta* [an interesting woman], *dominującą prezydent* [a dominating female president] (or, innovatively, *prezydentka* [a female president]);
— a pronoun/possessive: *jej praca* [her work], *tamtę osadniczki* [those female settlers], *ową nauczycielka* [that female teacher];
— a verb: *tak zadecydowałaś* [this is what you decided (spoken to a woman)], *one przywoływały* [they summoned (about women)], *czytelniczka napisała* [a female reader wrote], *Lila zgłosiła* [Lila (a female name) reported].

That way of introducing information about gender is particularly valuable, when a given masculine personal noun does not lend itself to creating a feminine personal form that would gain universal approval. For example, standard Polish widely accepts the structure like

Lekarz prowadząca mi zaleciła [The lead doctor prescribed (it) to me],

in which there is no agreement between the subject and the predicate (*lekarz zaleciła*) and between the subject and its designation (*lekarz prowadząca*). Here, the noun is used in generic reference, and the morpheme of the participle and the verb function as exponents of information about femininity — in both cases, it is the inflectional ending *-a*. On the other hand, if we decide to derive a

feminine personal noun from the masculine (*lekarka* [a female doctor]), the syntactic agreement will be preserved, but the utterance will be marked as colloquial:

> *Lekarka prowadząca mi zaleciła* [The female lead doctor prescribed (it) to me].

With reference to a man, the corresponding utterance has the form:

> *Lekarz prowadzący mi zalecił* [The lead doctor prescribed (it) to me].

Here, the noun *lekarz* and the participle *prowadzący* [lead] that describes it are in agreement, typical of Polish syntax, just like in the structure

> *lekarz zalecił* [the doctor prescribed].

Sometimes also the failure to inflect a noun indicates femininity.

> *Spotkam się z <u>lekarz</u>* [I will see a female doctor] (not-inflected) vs.
> *Spotkam się z <u>lekarką</u>* [I will see a female doctor] (inflected, with suffixal derivation).

> *Wywiad z <u>minister</u>* [An interview with a female minister] (not inflected) vs.
> *Wywiad z <u>ministrą/ministerką</u>* [An interview with a female minister] (*ministra* and *ministerka* are forms proposed by pro-gender-equality circles and currently not in universal use).

As shown above, a user of the Polish language has at their disposal a plethora of morphological means that expose information about femininity.

3.   From the 19th-century emancipatory breakthrough to the second decade of the 21st century

Today, the most heated discussions centre around word-formation techniques: the (in)ability to derive feminine personal nouns. Researchers are pointing out that the strategy for the derivation of feminine personal nouns has been productive for centuries (Klemensiewicz 1957: 106, 102; Kreja 1964: 129; Łaziński 2006: 247-248; Dembska 2012: 21-22), but its productivity was boosted after 1989 thanks to ideas and actions of the feminist and pro-gender-equality circles, who propose that feminine personal forms be derived from each masculine one. Historically, its creations of female names led to fewer normative doubts at the turn of the 20th century than they do today (cf. Woźniak 2014; Karamańska and Młynarczyk 2019; Małocha-Krupa 2018a).

When looking at the history of functioning of the feminine personal noun category, it is noticeable that its products appear in texts with a changing frequency, dependent on the ideology prevailing at the time, the adopted (or imposed) system of values, typical or characteristic of a given communicative community. Sometimes feminine personal nouns are omitted in communica-

tion (*invisibility of women*), while on other occasions their presence in the discourse is marked by high frequency. It happens when, for some reason or another, the gender parameter becomes communicatively relevant, in many cases, from the linguistic perspective, approaching textual redundancy (that excess of information about gender is one of the determinants of contemporary texts of pro-gender-equality circles). That is because feminine personal nouns belong to a word-formation category whose products are clearly and in many respects dependent on political and cultural factors. The increased productivity of the formations in question can be associated with the prevailing sociocultural trends and ideologies: once the emancipation of women, after WWII — social realism, and after 1989 — feminism (exemplification and analysis of the phenomenon are presented in the monograph: Małocha-Krupa 2018a).

3.1 The history of the functioning of feminine morphological structures can be divided into three periods. From the 2nd half of the 19th century, **emancipatory ideas**, contributing to women's increased presence in public and professional life, affected derivation of feminine personal noun innovations. Material excerpted from the Lviv-based fortnightly magazine 'Ster' (the editions from the years 1895–1897) indicates that the biggest number of new feminine personal forms belonged to two lexico-semantic fields. Those were names of positions of power (60% of the new language units, e.g. *inspektorka* [a female inspector], *ordynatorka* [a female department head]), names of professions and functions requiring tertiary education (59% of the new language units, e.g. *architektka* [a female architect], *weterynarka* [a female veterinary surgeon]). The names of the professions that women entered which came into use after the emancipation of women had started have been raising stylistic and word-formation doubts more often than those of the occupations and professions that women had traditionally followed (the lexical innovation rate of the latter amounts to 0%).

Analysis of the texts from the fortnightly 'Ster' reveals certain analogies with contemporary phenomena similar to the ideas of the gender equality discourse. It can be concluded on the basis of how nouns are worded in respect of the gender category (e.g. *właściciel/właścicielka* [a male/female owner], *czytelnik/czytelniczka* [a male/female reader], *wirtuoz/wirtuozka* [a male/female virtuoso], *wystawca/wystawczyni* [a male/female exhibitor]) that the communicating parties felt there was a need to differentiate the two gender categories, and quite often feminine personal forms were derived from masculine ones. Another manifestation of the productivity of the word-formation category in question was the existence of variants within it — linguistic formations performing a similar function in a given period of time and competing with each other for the same position within the language system. The Polish language of the second half of the 19th century had many instances of semantic doubles referring to names of women's professions (e.g. *pończoszniczka/pończoszarka* [a female stocking maker], *krawcowa/krawczyni* [a female dressmaker], *skrzypaczka/*

skrzypicielka [a female violin player]).

Discussions about the stylistic value of many names should be regarded as the first outcome of the achievements of women's emancipation (this is attested to by the earliest linguistic periodicals: *Poradnik Językowy* and *Język Polski*). Women, having gained election rights in 1918, in the newly reborn Poland, freed from the rule by the three partitioning countries, began to demand appellation corresponding to that of men. Gradually, they started to reject possessive suffixes, traditionally combined with surnames of maidens and wives (traditionally: *pan Małocha* [Mr Małocha], his wife→*Małosz-yna*, his daughter→*Małosz-anka*; *pan Nowak* [Mr Nowak], his wife→*Nowak-owa*, his daughter→*Nowak-ówna*), and, consequently, seeking the same designation as that of men (today only: *pani Małocha* [Ms Małocha], *pani Nowak* [Ms Nowak]), also as regards names of professions, which led to a weakening of the traditional derivational models and the strengthening of the trend to use some names of professions and academic titles generically. The published language use recommendations emphasising the need to accept and follow the regular word-formation patterns in Polish by deriving feminine personal nouns from masculine bases also evolved in the first half of the 20th century. When Poland regained its independence, the direct threat to the existence of its language and culture disappeared, and a relaxation, even if a partial one, of the language–tradition relationship and an opening to new communicative trends became possible. The fight to preserve the language thanks to its strict adherence to tradition was no longer the fundamental normative criterion. Increasingly, linguists started to support gender-neutral uses that were becoming more and more popular, instead of traditional solutions consisting in deriving feminine personal designations.

3.2 After WWII, the richness of the feminine lexis was a result of conscious propagandistic efforts by the authoritarian state, in which the equality of men and women was one of the slogans of Marxist-Leninist ideology. The new economic and political reality, the Polish People's Republic, demanded a new model of femininity. It was popularised by means of poster images, magazine covers, newsreels, one-off prints, literature and art. That 'poster image of a woman', saturated with ideological motifs, calling for taking on any social challenge, promoting equal access to work and women's professional mobilisation, often in jobs which were traditionally regarded as manly was accompanied by specific linguistic activities. Language was turned into a carrier, transmitter and cumulator of pragmatic values, propagated and controlled by the centralised communist apparatus that existed in the Polish People's Republic. To that aim, new feminine personal forms were derived, including *gwinciarka* [a female thread maker], *hutniczka* [a female steelworker], *instalatorka* [a female installer], *lutowaczka* [a female solderer], *spawaczka* [a female welder], which I refer to as propagandistic labels of socialist feminism. Their feminine personal grammatical structure was meant to persuade women (the persuasive function) that because the authorities were addressing them directly, they were becom-

ing important participants in civic life. The introduction of feminine personal nouns (to replace previous generic appellations) in the state-monopolised mass media was also intended to make it easier for women to identify with the jobs and positions the authorities were actively promoting (the identifying function). Being an expression of designation of the new, indoctrinated work reality, desired by the state-building apparatus, the feminine derivates validated its authenticity (the magical function). The above conclusions are based on an analysis of a corpus of texts from some newspapers and magazines published in the years 1946–1956 (primarily 'Przyjaciółka', 'Trybuna Ludu', 'Kraj', 'Mucha', 'Moda i Życie', 'Naprzód Dolnośląski' and 'Żołnierz Polski') that I have collected.

The belief in the equality of men and women, grounded in the social awareness of the Polish People's Republic, was understood as 'one and the same', i.e. 'being the same under law and at work', in line with the propagandist and constitutional image of emancipation of women. For that reason, the professional appellations of men and women were gradually becoming uniform and assuming an identical, masculine form, connoting appropriate prestige, formality, domination, higher rank and normativity. The word-formation indicators of femininity were disappearing from communication, while at the same time the trend to use the same names to refer to both genders strengthened. This conclusion particularly relates to formations denoting a certain social hierarchy (names of positions of authority, such as *ambasador* [ambassador], *generał* [general], *kanclerz* [chancellor], *marszałek* [marshal], *minister* [minister], *premier* [prime minister], *prezydent* [president]) as well as occupations or statuses recognised as prestigious (*adwokat* [advocate], *prokurator* [attorney], *prezes* [chairperson]).

3.3 Because of the trend that grew in the subsequent decades of the 20th century towards generic usages and in the light of a declining productivity of the feminine personal noun category, after the political and cultural breakthrough of 1989 (a redevelopment of the political system, the first free, democratic elections to the Polish parliament), the feminist and leftist circles started to promote the idea of linguistic-gender symmetry and of deriving a semantically equivalent feminine personal form from each masculine one. This idea was and still is informed by the belief that it is worth extracting from the linguistic void something that which is socially relevant, which explicates the gender of communicating parties. And this factor in connection with postmodern constructs, identity attitudes and anti-discriminatory concepts, has become a communicationally pertinent parameter and a relevant element of the new, equality-conscious language etiquette (Małocha-Krupa, Hołojda, Krysiak and Pietrzak 2013). The feminist theories (especially those inspired by the second wave of feminism), both those literary-science-oriented and those within the area of gender linguistics, focus, among other things, on the connection between the symbolic (linguistic) and social orders. They regard language and communication as sig-

nificant indicators of patriarchal attitudes, tools for gender-based distribution of social roles and statuses, reservoirs of ways of perception and categorisation of information about gender, which — in this study perspective — appear as asymmetrical and hierarchical. They also indicate that all masculine is privileged at the expense of the feminine — subordinated to masculine notions, ways of perception, designation and valuation. Such a critical analysis of linguistic and discursive reality is related to a more or less directly explicated reformative approach to manifestations of the privileged position of either gender in the symbolic order, its exclusion or depreciation. Thus, the study of gender linguistics and feminist criticism is of a teleological nature — through description, analysis, and subsequent reinterpretation of areas that require change, they attempt to deconstruct the traditional conceptual or grammatical system and to establish a new one, based on the concept of equality. The above-mentioned analyses and communicative practices aim to promote the 'coming out of silence' (primarily by women), the naming of the unnamed, the finding of one's own voice, and lexical and semantic symmetry in designating men and women. Projects developed from the feminist perspective, aimed at balancing social relations through making changes to language also left their mark on the Polish language and have been influencing its development for nearly three decades. The ideas of gender linguistics and feminist criticism of language initiated *sui generis* activation of the lexical and word-formation layers of Polish which relate to nominative operations on the feminine category. Most probably the history of productivity of feminine personal nouns would have been different if not for the involvement of feminist and pro-equality circles in propagating the idea to complement the existing lexical and semantic gaps as regards designating women, particularly their professions, positions and social functions. That is probably why there is a widespread belief that 'feminine personal nouns have been invented by feminists', 'bored female academicians' — however, as analysis has shown: Kreja 1964; Woźniak 2014; Małocha-Krupa 2018a; Karamańska and Młynarczyk 2019; Młynarczyk 2019, such nouns had been an important part of the word-formation system of the Polish language earlier. As Prof. Jan Miodek has been saying during his lectures at Wrocław University for the last few decades: 'Poles' natural word-formation reflex means creation of feminine names'. Feminist movements have only managed to restore a portion of the traditional feminine personal lexis, subjecting some of the old names to neosemantisation, melioration of their meaning (e.g. *adwokatka* [a female advocate], *mecenaska* [a female lawyer], *ministerka* [a female minister]), and setting in motion derivation of innovative formations, required to meet the new naming needs with a view to achieving the equality vision between the symbolic linguistic reality and the extra-linguistic one. The return of feminine personal nouns to general Polish, especially that used in journalism, is an ongoing process. We have observed a stylistic reclassification of many a formation that was originally used in a limited environment or was perceived as marked in general Polish. However, numerous feminine personal nouns, especially those

naming degrees, professions or statuses characterised by high social prestige, are regarded by a majority of speakers as stylistically non-neutral. They are thus not accepted, and are often perceived as reflections of feminist or pro-gender-equality views. The nouns are again associated with ideology, although they are also often alleged to sound depreciatory or humorous.

4.   Codification of feminine person nouns as a challenge to lexicography

Today, we are seeing a revival of feminine personal nouns (feminine, suffixal derivation) in the Polish language used by pro-gender-equality circles, efforts to expand the limits of normativeness, the overcoming of the limits of codification, which are affecting the condition of the general Polish language. Usages with both-gender references are competing with increasingly more frequent feminine personal noun updates. For years we have seen discussions on their functionality, acceptability and stylistic value (cf. e.g.: Łaziński 2006, Hołojda 2013, Kiełkiewicz-Janowiak 2019, Szpyra-Kozłowska 2019a, b) as well as the ability to examine them in detail. As one of the researchers has noticed:

> In a country with rather conservative attitudes, both to gender roles and language use norms, it might be sensible to dissociate the issue from the political agenda it is in fact part of. (Kiełkiewicz-Janowiak 2019: 167)

The issue regarding feminine personal nouns is neither new nor one-dimensional. Undoubtedly, dictionaries of general Polish have failed so far to rise to the challenge to provide a fuller, objective codification of the language units in question. Their position in contemporary lexicography is 'indefinite, uncertain' (Krysiak 2016: 84). Dictionary authors either ignore them or do not agree as to their stylistic value, assigning a stylistic qualifier to them, or the stylistic value assigned by the authors is not accepted by all parties to communication. Patrycja Krysiak, author of, for instance, a monograph *Feminine Names in Contemporary Polish and French Lexicography. Comparative Analysis of Selected Most Recent Polish and French General Dictionaries* (2020), already noticed in one of her earlier publications that: 'exclusion of feminine personal nouns by lexicographers is causing language users to treat them with a considerable degree of uncertainty — after all, a dictionary is the basic source of knowledge about language for non-specialists' (Krysiak 2016: 89). On the one hand, they often assume that if a given word is not to be found in a dictionary, it should not be used in communication, and, on the other hand, conscious communication participants are aware of the need to adjust their lexical choices to the principles of communication that are changing right in front of us, particularly as regards the gender-equality etiquette. There is also a growing demand for explicit knowledge, recipe-like knowledge regarding which words should be used in the public space and how, and this demand is enormous. This is attested by questions sent to language advisory services and numerous emails asking for resolving normative doubts that I receive — they might be used as a basis for a language guide on

how to communicate gender-equality today.

Consequently, lexicographers are faced with an uneasy task. They recognise the dynamism of change in the habits regarding communication about gender, the social demand for their description and codification. However, the lexical material relating to the phenomenon is extremely difficult to describe unambiguously and in a simple way. At the present stage of changes in the normative and gender-equality awareness, in the early 2020s, one should even ask whether it is at all possible to codify and stylistically qualify women's designations in a way that would be fully scientific, objective, free from any individual or subjective opinions. I believe not. Any specific form is only included in codification if someone decides to do so, using of course some previously adopted criteria.

This is attested to by discussions on stylistic value of numerous feminine personal nouns: names of positions of power (*dyrektorka* [a female director], *ministerka* [a female minister], *prezydentka* [a female president]), names of professions (*nauczycielka* [a female teacher], *prawniczka* [a female lawyer], *psycholożka* [a female psychologist]), and characterising names (*interpretatorka* [a female interpretator], *przodkini* [a female ancestor], *gościni* [a female guest]). Their use and valuation depends on many parameters: the communicative situation, an attitude to language, views on social reality, sensitivity to gender-equality issues, as well as aesthetic sense and associations that a given form triggers.

5.   A dictionary of female nouns

The doubts referred to above and lexicographers' inconsistency in recording (or omission of) feminine personal forms led to the idea to publish a 'Dictionary of Polish Female Nouns'. Several female linguists from Wrocław University, Poland: Katarzyna Hołojda-Mikulska, Patrycja Krysiak, Marta Śleziak, the author of this article, Agnieszka Małocha-Krupa, undertook a pioneering attempt to scientifically tackle the issue, and published in 2015 *Słownik nazw żeńskich polszczyzny* [Dictionary of Polish Female Nouns; further on referred to as SNŻP].

The concept of the dictionary relates to 'the typological truth about the centuries-long richness of the Polish language in terms of derivation processes within the category of femininity' (Miodek 2017: 172). It contains feminine lexemes from texts written during the period spanning the second half of the 19th century and the present times.

This is the first dictionary in Polish lexicography to only contain feminine structures, in linguistics referred to as feminine personal nouns (cf. Sec. 1). Before, feminine personal nouns were made part of other publications, found their way into language use guides and normative books, for instance works devoted to word-formation trends in Polish. However, there had been no special publication recording authentic uses of feminine personal nouns in Polish,

excerpted from written texts.

The total number of the incorporated lexical items amounts to 2,103, including 422 (i.e. 20.07%) that until 2014 had not appeared in any dictionaries of general Polish language (that group has both some innovative formations, like *bodypainterka* [a female bodypainter], *bookcrosserka* [a female bookcrosser], *brafitterka* [a female bra-fitter], *castingowiczka* [a female participant in a casting], *copywriterka* [a female copywriter], *forumowiczka* [a female forum participant], *ghostwriterka* [a female ghostwriter], *headhunterka* [a female headhunter], *japiszonka* [a female yuppie], *lobbystka* [a female lobbyist], *performerka* [a female performer], *researcherka* [a female researcher], *senselierka* [a female smell expert], *shopperka* [a female shopper], *slamerka* [a female participant in a poetry slam], *streetworkerka* [a female streetworker], *trendsetterka* [a female trendsetter], *vlogerka* [a female vlogger], and some traditional ones, which for various reasons had not been codified in other dictionaries, e.g. *eksperymentatorka* [a female experimenter, *ratowniczka* [a female rescuer], *ubezpieczycielka* [a female insurance agent]). SNŻP is, above all, a source of knowledge about language and its transformations related to changing reality. It presents feminine personal nouns used in public discourse: old ones and nearly forgotten ones (*świekra* [the husband's mother], *zełwa* [the husband's sister]) as well as some entirely new ones (*couchsurferka* [a female participant in couchsurfing], *galerianka* [a young female student offering sexual services at a shopping centre]); referring to professions that are rare or non-existent today (*pończoszniczka* [a female stocking maker], *sekserka* [a female that determines sex of chicks]) or those that have appeared recently (*bodypainterka* [a female bodypainter], *profilerka* [a female profiler]); referring to women living in the peculiar reality of the Polish People's Republic (*formiarka* [a female form maker], *traktorzystka* [a female tractor driver]) or to social roles that never change (*matka* [a mother], *żona* [a wife]).

The publication constitutes a scientific description of a section of reality that is reflected in language. It is based on the principle of an objective search for and recording of language forms that fulfil appropriate criteria, and the editing approach was to adopt a method of description of the gathered material that would be free from any direct valuation of each entry. The over-riding goal was to show the richness of feminine personal nouns, their diversity, variable stylistic value that sometimes comes to light, particularly well visible in the case of nouns attested to with 19th-century uses.

The dictionary may be treated as an open file and a collection of word forms to choose from when composing one's own texts, adopting a traditional approach or a pro-gender-equality one towards the changing social reality, open to linguistic innovation or making use of the mine of nouns, often forgotten ones. It can also be used for checking usage contexts and the presence of feminine personal nouns in official texts. The book can also be looked upon as a guide to a world in which women play an important role, and its reader may find out, from the carefully-considered selection of quotations, how their social roles have changed, how the professional and family life they create has evolved,

and how our reality (not only that of women) has changed over the years.

5.1   Some of the lexicographical assumptions

Słownik nazw żeńskich polszczyzny uses an innovative lexicographical method, particularly as regards the layout of entries and introduction of direct and indirect description qualifiers. It was assumed that the quotation, text and discourse would play the primary role; therefore, the approach was called a discourse-centred method. In it, a lexeme is directly highlighted by a quotation, and a quotation by a comment, so the definition of the lexeme is given at the end — sometimes with synonyms as cross-references to other entries in the collection of texts. The definition constitutes the last structural element of an entry. This solution does away with the traditional order and the linearity of the sequential, expected, predictable elements of a typical microstructure of an entry, but it is conducive to involving the reader in the text and making them understood on their own the meaning of the lexeme and how it is used on the basis of its textual representation.
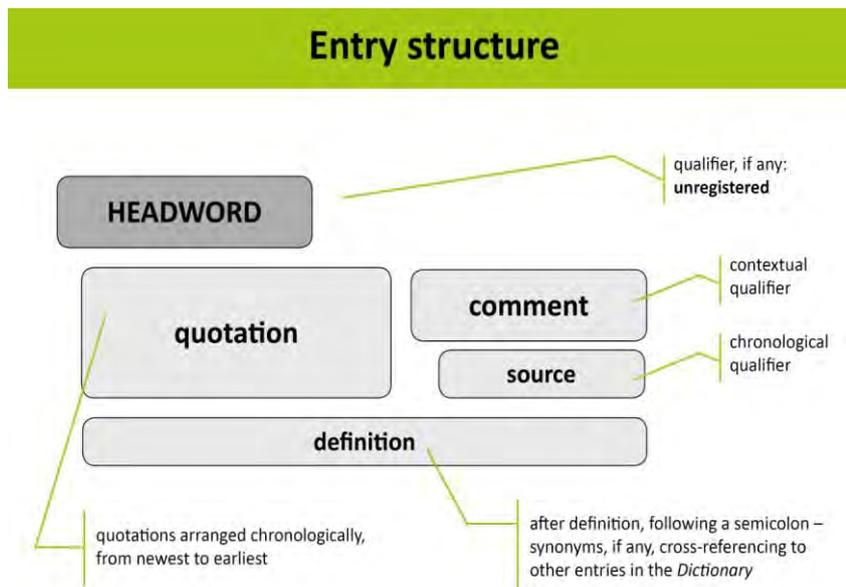
This concept was appreciated by the scientific reviewers of the dictionary. Prof. dr hab. Mirosław Bańko wrote:

> I also support the idea of the dictionary speaking to the reader above all in the voices of the quoted authors and authoresses, often well-known public figures, either living now or a hundred years ago. Thanks to the quotations, the book

gains credibility and historical depth, and involves the reader intellectually as well as emotionally.

The method adopted is described both in the *Introduction* to the dictionary, and in some published articles: Śleziak 2018; Małocha-Krupa 2018b.

The publication constitutes a scientific description of a section of reality that is reflected in language. Because of the adopted method (a descriptive, not a prescriptive dictionary), it documents the variety and productivity of feminine personal noun formation.

Below — illustratively — I present two examples of entry-level articles from SNŻP:

A

aptekarka

[…] rozumne garnęły się skwapliwie do doktorek, aptekarek, dentystek, darzyły je zaufaniem i brały niejako udział w ich pochodzeniu, rozumiejąc doniosłość sprawy, jakiej były przedstawicielkami.

Pisała w 1903 roku Waleria Marrené z Malletskich (pisarka, tłumaczka) w tekście publicystycznym pt. „Kobieta czasów współczesnych"; źródło: Antologia 1999, s. 86

Kobiety w aptekarskim zawodzie w ogóle z pożytkiem pracować nie mogą: w zamian nie mamy nic przeciwko zatrudnianiu ich, jako uczennic i aptekarek, w aptekach szpitalnych, w lecznicach itp. instytucjach. Co na to powiedzą aptekarki belgijskie i angielskie, a no i nasze także?

Wypowiedź niemieckiego referenta Rady aptekarskiej; źródło: „Ster" 12/1897, s. 190

[…] uzupełniła ten wniosek żądaniem, aby w tym jury zasiadły także i kobiety aptekarki.

Z dyskusji na kongresie farmaceutycznym w Mons, w Belgii; źródło: „Ster" 4/1897, s. 62

~ pracowniczka lub właścicielka apteki; *farmaceutka*

aptekarka [female pharmacist]

[...] the sagacious ones would come eagerly to female doctors, pharmacists and dentists, they trusted them and somehow participated in their background, appreciating the significance of the matter that they represented.

Wrote (in 1902) Waleria Marrené of the Malletski family (a writress and translator) in an article entitled 'Kobieta czasów współczesnych' [The woman of the modern age]; source: Antologia 1999, p. 86

Women cannot be of use in pharmacy business; instead, we do not mind if they are employed as female trainees or pharmacists at hospital pharmacies, infirmaries or similar institutions. What will Belgian and English, as well as our, female pharmacists say to that?

Said a German clerk from the Pharmacist Council; source: 'Ster' 12/1897, p. 190

[...] she pointed out in her request that the jury should also include female pharmacists.

From a discussion at a pharmacist congress in Mons, Belgium; source: 'Ster' 4/1897, p. 62

~ a female employee or owner of a pharmacy; *farmaceutka*

F

fachowczyni *nierejestrowany*

Od zarania skrzypaczka, po wszystkich wymaganych szkołach: SP Muz, GIM Muz, AM Muz; lata doświadczeń zdobytych w Schola Cantorum et Musicorum Cathedralis Sttetinensis. Fachowczyni w dziedzinie muzyki liturgicznej. Aktualnie nauczycielka skrzypiec.

O organistce parafii p.w. Przemienienia Pańskiego w Załomiu; źródło: parafiazalom.pl, dostęp: 28 XII 2013

Potrzebę łączności zawodowej odczuwały wszystkie zgromadzone fachowczynie […].

Sprawozdanie z działalności Delegacji Kobiet w Warszawie w 1894 roku; źródło: "Ster" 1/1895, s. 14

~ specjalistka w jakimś fachu, ta, która dobrze się zna na swojej pracy; *ekspertka; profesjonalistka; specjalistka; znawczyni*

fachowczyni [female specialist] *unregistered*

Since always — a female violinist, with all the required schools completed: primary school of music, middle school of music, university of music; years' worth of experience at Schola Cantorum et Musicorum Cathedralis Sttetinensis. A female specialist in liturgical music. Currently, a female violin teacher.

About the female organ player of the parish of the Transfiguration of Jesus at Załom; source: parafiazalom.pl, access: 28 Dec 2013

All of the gathered female specialists felt the need for professional connections [...].

A report on the operation of the Women's Labour Committee in Warsaw in 1894; source: 'Ster' 1/1895, p. 14

~ a female specialist in some field; a woman expert in her profession; *ekspertka, profesjonalistka, specjalistka, znawczyni*

5.2   Consequences of descriptivism

The dictionary's descriptiveness results not only from the democratic view of the language that is close to its editors but also from the fact that in many cases at the present stage of the development of the Polish language, when various ways of introducing information about the female gender into text are competing with one another (cf. Section 2) and parties to communication have varying approaches and axiological systems (cf. Section 4), the lexicographer who is not always objective is unable to unambiguously and accurately establish the stylistic value of a given feminine personal noun. In the case of nouns that are currently stylistically suspicious to some users this seems impossible. Frequently, in the same contexts, text types/genres and discourses, both a generic and a feminine personal form can be used.

Many people will regard feminine personal nouns such as *kierowniczka* [a female manager] or *dyrektorka* [a female director] as non-neutral stylistically

and will qualify them as colloquial, while other people, not always those in favour of gender equality, will use such words to introduce themselves in formal situations, will use them in writing, put on business cards or websites of the institution they represent (i.e. in official language contact situations). In the normative awareness of such people, such forms are neutral stylistically.

That is why in the case of SNŻP the decision was made to do away with — as worded by Marta Śleziak (2018: 254) 'arbitrary normative solutions'. As far as stylistically doubtful forms are concerned, the reader will have to choose the form themselves — either the suffixal one, which they will find in the dictionary (e.g. *analityczka* [a female analyst], *dyrektorka* [a female director], *prezeska* [a female president] — stylistically doubtful to some people), or the both-gender variant, which — in line with the adopted assumptions — are not included in the dictionary. Plenty of stylistic information can be derived from the material documentation (quotations) provided by the authors on purpose, as well as commented contexts in which a given feminine personal noun is used and its sources (Śleziak 2018).

6.   In conclusion

The publication of SNŻP confirmed the existence of a significant social demand for linguistic opinions on the designations of women as regards their occupations, positions, public functions and academic degrees. Two days after the meet-the-authors session the dictionary was out of print, and the reprint was sold out within several weeks. Dozens of invitations to give lectures, numerous telephone calls and e-mails requesting interviews, as well as a constant flow of requests for expert advice (primarily by e-mail) on feminine personal nouns, their creation and stylistic value attest to the lively reception of the publication.

Those who have had a careful look at SNŻP (*Przegląd Uniwersytecki* 2016(2)) have recognised the values of the dictionary, e.g. that arising from the fact that 'although it focuses on a selected, single word-formation and semantic category, i.e. female personal nouns, it shows their broad textual, pragmatic and sociocultural background'. (Rzymowska and Poprawa 2016: 36), The material collected in it enables you to track the paths of the history of the feminine personal noun category — its disappearances and reappearances in Polish. The publication's reviewers have also expressed their appreciation for the adopted word-formation techniques, thanks to which — in their opinion — the SNŻP collection of texts can be an inspiration and an object of further lexical and semantic studies, both into feminine personal nouns functioning in official language, and into neologisms and nonce words reflecting current cultural processes.

The dictionary's authors are now working on a new, expanded edition of SNŻP, which will introduce a modified entry structure.

References

Dembska, K. 2012. *Tendencje rozwojowe polskich i rosyjskich nazw zawodowych kobiet na tle języka czeskiego* [The Development Tendencies in Polish and Russian Female Profession Names within the Context of the Czech Language]. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Hołojda, K. 2013. Jak Polki postrzegają feminatywy? [What Do Polish Women Think about Feminine Personal Nouns?] Małocha-Krupa, A., K. Hołojda, P. Krysiak and W. Pietrzak. 2013. *Równościowy savoir-vivre w tekstach publicznych* [Gender-equality savoir-vivre in Public Texts]: 101-105. Warsaw: The Chancellery of the Prime Minister, Office of the Government Representative for Equal Treatment.
http://www.ifp.uni.wroc.pl/data/files/pub-9470.pdf

Karamańska, M. and E. Młynarczyk. 2019. *Komponenty eksponujące cechę żeńskości w nazwach stowarzyszeń II Rzeczpospolitej* [Exponents of the Feature of Femininity in Names of Associations in the Second Republic of Poland]. Annales Universitatis Paedagogicae Cracoviensis 283. *Studia Linguistica* 14: 67-78.

Karwatowska, M. and J. Szpyra-Kozłowska. 2005. *Lingwistyka płci. Ona i on w języku polskim* [Gender Linguistics. She and He in the Polish Language]. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Kiełkiewicz-Janowiak, A. 2019. Gender Specification of Polish Nouns Naming People: Language System and Public Debate Arguments. *Slovenščina 2.0: Empirical, Applied and Interdisciplinary Research* 7(2): 141-171.
https://doi.org/10.4312/slo2.0.2019.2.141-171

Klemensiewicz, Z. 1957. *Tytuły i nazwy zawodowe kobiet w świetle teorii i praktyki* [Women's Professional Titles and Names in Theory and Practice]. *Język Polski* 37(2): 101-119.

Kreja, B. 1964. Słowotwórstwo nazw żeńskich we współczesnym języku polskim (wybrane zagadnienia) [Word Formation of Polish Female Personal Nouns (Selected Issues)]. *Język Polski* 44(3): 129-140.

Krysiak, P. 2016. Feminatywa w polskiej tradycji leksykograficznej [Feminine Personal Nouns in Polish Lexicographical Tradition]. *Rozprawy Komisji Językowej WTN* [Papers of the Language Committee of Wrocław Scientific Association]. Wrocławskie Towarzystwo Naukowe, Yearbook 42: 83-90.

Krysiak, P. 2020. *Nazwy żeńskie we współczesnej leksykografii polskiej i francuskiej. Analiza porównawcza wybranych najnowszych polskich i francuskich słowników ogólnych* [Feminine Names in Contemporary Polish and French Lexicography. Comparative Analysis of Selected Most Recent Polish and French General Dictionaries]. Wrocław: Oficyna Wydawnicza ATUT.

Łaziński, M. 2006. *O panach i paniach. Polskie rzeczowniki tytularne i ich asymetria rodzajowo-płciowa* [On Men and Women. Polish Titular Nouns and Their Sex/Gender Asymmetry]. Warsaw: Wydawnictwo Naukowe PWN.

Małocha-Krupa, A. (Ed.). 2015. *Słownik nazw żeńskich polszczyzny* [Dictionary of Polish Female Nouns]. Entries by: Hołojda K., P. Krysiak, A. Małocha-Krupa and M. Śleziak. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Małocha-Krupa, A. 2018a. *Feminatywum w uwikłaniach językowo-kulturowych* [Feminitives and Their Linguistic and Cultural Entanglements]. Wrocław: Oficyna Wydawnicza ATUT.

Małocha-Krupa, A. 2018b. Opis leksykograficzny feminatywum. (Nie)możliwości zobiektywizowania [Lexicographic Description of Feminine Personal Nouns. (In)ability to Objectivise]. Bańko, M. and H. Karaś (Eds.). 2018. *Między teorią a praktyką. Metody współczesnej leksykografii* [Between Theory and Practice. Methods in Contemporary Lexicography]: 151-165. Warsaw: Wydawnictwo Uniwersytetu Warszawskiego.

Małocha-Krupa A., K. Hołojda, P. Krysiak and W. Pietrzak. 2013. *Równościowy savoir-vivre w tekstach publicznych* [Gender-equality savoir-vivre in Public Texts]. Warsaw: The Chancellery of the Prime Minister, Office of the Government Representative for Equal Treatment. http://www.ifp.uni.wroc.pl/data/files/pub-9470.pdf

Miodek, J. 2017. *Słownik nazw żeńskich polszczyzny* [recenzja]. Dictionary of Polish Female Nouns [A Review]. Sojka-Masztalerz, H., M. Porawa and T. Piekot (Ed.). 2017. *Obraz wojny w mediach, pamięci i języku* [The Picture of War in the Media, Memory and Language]: 171-172. Oblicza Komunikacji [Facets of Communication] 10(2017). Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Młynarczyk, E. 2019. Tendencje nazwotwórcze w chrematonimii społecznościowej II Rzeczypospolitej (na przykładzie nazw stowarzyszeń) [Trends Related to Creating Names in Social Chrematonymy of the Second Polish Republic (Based on the Names of Associations)]. *Onomastica* 63: 209-226.
https://onomastica.ijp.pan.pl/index.php/ONOM/article/view/78

Nowosad-Bakalarczyk, M. 2009. *Płeć a rodzaj gramatyczny we współczesnej polszczyźnie* [Natural and Grammatical Gender in Contemporary Polish]. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Rzymowska, L. and M. Poprawa. 2016. Portret kobiety w słowach zapisany [A Portrait of a Woman as Recorded in Words]. *Przegląd Uniwersytecki* 2016(2): 36-37.
https://www.bu.uni.wroc.pl/sites/default/files/images/doc/PU_2016_02.pdf

Szpyra-Kozłowska, J. 2019a. "Premiera", "premierka" czy "pani premier"? Nowe nazwy żeńskie i ograniczenia w ich tworzeniu w świetle badania ankietowego ['Premiera', 'premierka' or 'pani premier'? New Female Personal Nouns and Constraints on Their Formation in the Light of a Questionnaire Study]. *Język Polski* 99(2): 22-40.
https://www.ceeol.com/search/article-detail?id=780107

Szpyra-Kozłowska, J. 2019b. Feminitives in Polish: A Study in Linguistic Creativity and Tolerance. Bondaruk, A. and K. Jaskuła (Eds.). 2019. *All Around the Word: Papers in Honour of Bogdan Szymanek on his 65th Birthday:* 339-364. Lublin: Wydawnictwo KUL.

Śleziak, M. 2018. Prymat leksykograficznego egzemplum i jego konsekwencje dla hasła — na przykładzie Słownika nazw żeńskich polszczyzny [Primacy of Lexicographical Example and Its Consequences for the Entries — On the Example of the Dictionary of Polish Feminatives]. Bańko, M. and H. Karaś (Eds.). 2018. *Między teorią a praktyką. Metody współczesnej leksykografii* [Between Theory and Practice. Methods in Contemporary Lexicography]: 245-255. Warsaw: Wydawnictwo Uniwersytetu Warszawskiego.

Woźniak, E. 2014. Język a emancypacja, feminizm, gender [Language and Emancipation of Women, Feminism, Gender]. *Rozprawy Komisji Językowej Łódzkiego Towarzystwa Naukowego* [Papers of the Language Committee of Łódź Scientific Association] 60: 295-312.
http://www.rjp.pan.pl/index.php

Das Modul *Werbeslogans*. Eine korpusinformierte lexikografische Ressource zum aktuellen Gebrauch von Werbeslogans außerhalb der Domäne Werbung

Janja Polajnar, *Department of German, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia (janja.polajnar@ff.uni-lj.si)*

Zusammenfassung: Der vorliegende Artikel setzt sich zum Ziel, das Modul *Werbeslogans*, das in das *OWID-Sprichwörterbuch* am Leibniz-Institut für Deutsche Sprache nachhaltig integriert ist, in seiner Entstehung darzustellen. Es handelt sich um eine korpusinformierte und nach Kriterien der wissenschaftlichen Lexikografie erarbeitete Onlinedokumentation von Werbeslogans bzw. von aktuell gebräuchlichen verfestigten Sätzen aus der Werbung, die bereits Einzug in die Gemeinsprache gefunden haben. Da diese Slogans ähnlich wie Sprichwörter gebraucht werden, wurden sie auch ähnlich beschrieben. Das Modul basiert auf den innovativen lexikografischen Konzepten im Sprichwortbereich, die im EU-Projekt *SprichWort. Eine Internetplattform für das Sprachenlernen* für die *Sprichwort-Plattform* entwickelt wurden, und auf dem *OWID-Sprichwörterbuch*; es ist innovativ und stellt keine Fortschreibung tradierter Wörterbücher im Bereich der Phraseologie dar. Im Artikel wird einerseits die korpusinformierte Methodik zur Analyse des Slogangebrauchs außerhalb der Domäne Werbung expliziert und andererseits der lexikografische Prozess, die Konzeption der Mikrostruktur der Slogan-Artikel sowie die verschiedenen Zugriffsmöglichkeiten in *OWID (Online-Wortschatz-Informationssystem Deutsch)* dargestellt. Die dargestellte lexikographische Behandlung von Werbeslogans im Rahmen von OWID ist (in der germanistischen Lexikographie) ein Novum und trägt zur lexikographischen Behandlung von polylexikalen Lexikon-Einheiten bedeutend bei; vor allem aus der Sicht der neueren Betrachtungen des Lexikons aus der Perspektive der Konstruktionsgrammatik.

Schlüsselwörter: MODUL *WERBESLOGANS*, ONLINE-SLOGAN-ARTIKEL, SLOGAN-GEBRAUCH, SLOGAN-VARIANTEN, SLOGAN-MUSTER, LEXIKOGRAFISCHE RESSOURCEN, KORPUSINFORMIERTE INTERNETLEXIKOGRAFIE, LEXIKOGRAFISCHE ONLINE-SLOGAN-ARTIKEL

Abstract: E-Module *Werbeslogans*. A Corpus Informed Lexicographic Ressource of Advertising Slogan Use Outside the Domain of Advertising. The article aims to describe the development of E-Module *Werbeslogans*, integrated into *OWID-Sprichwörterbuch* at the Leibniz Institute for German Language in Mannheim. It represents a corpus

informed online description and presentation of advertising slogans, i.e. set sentences from adver-
tising, currently used outside the domain of advertising in everyday language. As these advertising
slogans are used in the same way as proverbs, they can be similarly lexicographically described.
The module is based on the lexicographic concepts developed in the EU project *SprichWort. Eine
Internetplattform für das Sprachenlernen* für die *Sprichwort-Plattform* as well as the proverb articles in
the *OWID-Sprichwörterbuch*; it is innovative and does not represent a continuation of traditional
dictionaries in the field of phraseology. The article describes the corpus informed method for the
analysis of advertising slogans used outside of the domain advertising, continues with the descrip-
tion of the lexicographic process as well as the microstructure of online slogan articles and con-
cludes with the links that connect different information at the macrostructural level of *OWID
(Online-Wortschatz-Informationssystem Deutsch)*.

Keywords: MODULE *WERBESLOGANS*, ONLINE SLOGAN ARTICLES, ADVERTISING
SLOGAN USE, SLOGAN VARIANTS, SLOGAN PATTERNS, LEXICOGRAPHIC RESSOURCES,
CORPUS INFORMED INTERNETLEXICOGRAPHY, LEXICOGRAPHIC ONLINE-SLOGAN-
ARTICLES

1.   Einleitung

Heute steht einem eine Vielfalt an unterschiedlichen, online frei verfügbaren
lexikografischen Ressourcen in vielen Sprachen zur Verfügung, denn die lexi-
kografische Forschung und Praxis haben sich seit Mitte der 90er Jahre durch
das Internet, den Computereinsatz und die Untersuchungen umfangreicher
elektronischer Korpora[1] funktional und strukturell stark verändert und modi-
fiziert (Stichwort: Internetlexikografie) (vgl. Klosa und Müller-Spitzer 2016: XI).

> Wörterbuchportale, semi-automatisch generierte Angaben, komplexe Suchfunk-
> tionen und ein moderater Umfang an Multimodalität sind heute Usus. Auch die
> lexikographische Arbeitsumgebung spiegelt diese Entwicklung wider mit der
> Verbindung zwischen Korpora und intelligenten Tools zur Extraktion von Daten
> und verschiedenen Möglichkeiten der Verbindung zu texttechnologischen
> Anwendungen (Engelberg, Klosa-Kückelhaus und Müller-Spitzer 2019: 30).

Lexikografische Online-Ressourcen zur deutschen Sprache werden beispiels-
weise vom Dudenverlag (*Duden Onlinewörterbuch*), von der Union deutscher
Akademien (z. B. *DWDS, Wörterbuchnetz* u. a.) und auch am Leibniz-Institut für
Deutsche Sprache in Mannheim (*OWID*) (vgl. Klosa-Kückelhaus und Müller-
Spitzer 2019: 418f.) erstellt. Im Wörterbuchportal *OWID* (*Online-Wortschatz-
Informationssystem Deutsch*), das sich auf „Ressourcen zu spezialisierten Wort-
schatzbereichen konzentriert" (ebd.), stehen den User/-innen zehn lexikogra-
fische Ressourcen zur Verfügung (Stand März 2021)[2], darunter auch das
OWID-Sprichwörterbuch, eine korpusbasierte und „nach Kriterien der wissen-
schaftlichen Lexikografie erarbeitete Dokumentation aktuell gebräuchlicher
fester Sätze der deutschen Sprache — im Kern Sprichwörter" (*OWID-Sprich-
wörterbuch*)[3] (Abb. 1). Ins Sprichwörterbuch wurde zuletzt das Modul *Werbe-
slogans* (Steyer und Polajnar 2015) integriert. Hierbei handelt es sich um die

lexikografische Beschreibung des aktuellen Korpusgebrauchs von Werbeslogans, die in der Gemeinsprache eine gewisse Geläufigkeit aufweisen. Da diese Slogans ähnlich wie Sprichwörter funktionieren, wurden sie nach demselben Modell (Steyer 2012; Steyer und Ďurčo 2013; Ďurčo, Steyer und Hein 2017) mit einigen Änderungen in der Wörterbuchartikel-Struktur beschrieben und in das Sprichwörterbuch integriert. Die Slogan-Artikel zeigen nicht nur, wie lebendig und variabel Werbeslogans in außerwerblichen Kontexten sind und wozu Sprecher/-innen sie benutzen, sondern auch, dass die Werbesprache eine moderne Quelle für die Entstehung neuer Sprichwörter darstellt.
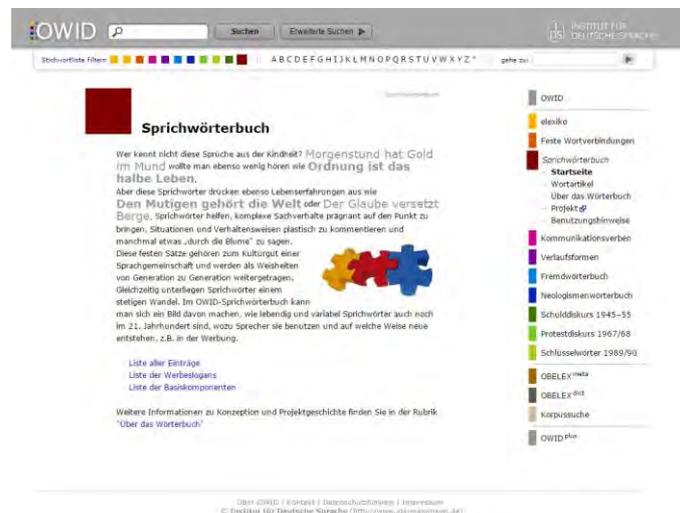
Abb. 1:   Das *OWID-Sprichwörterbuch* mit der *Liste der Werbeslogans* und weiteren lexikografischen Ressourcen des Wörterbuchportals *OWID* (rechte Spalte)

Dem Modul *Werbeslogans* liegt eine umfangreiche korpusinformierte Untersuchung zugrunde. Darin wurde der Slogangebrauch in Zeitungskorpora des Deutschen Referenzkorpus (vgl. *DeReKo*) im Zeitraum 1990–2008 diachron untersucht und die Mikrodiachronie ihres Vorkommens sowie die Tendenzen ihres Bedeutungswandels, ihrer Varianz und Musterbildung und ihrer kontextuellen Einbettung in außerwerblichen Kontexten nachgezeichnet (Polajnar 2012, 2016, 2019).

Das neue Modul ist ähnlich wie das Sprichwörterbuch selbst eine korpusinformierte und „nach Kriterien der wissenschaftlichen Lexikografie erarbeitete Onlinedokumentation aktuell gebräuchlicher verfestigter Sätze" (*OWID-Sprichwörterbuch*) aus der Werbung. Die Dokumentation von Werbeslogans wird mithilfe systematischer, empirischer Erhebungen mit COSMAS II (*DeReKo*) und anschließend mit *lexpan* (2017), dem Analysewerkzeug zur Systematisierung von KWIC-Zeilen, erarbeitet. Folglich wird nicht an die Fortschreibung tra-

dierter Wörterbücher angeknüpft. Das Modul basiert einerseits auf den lexiko-
grafischen Konzepten, die im EU-Projekt *SprichWort. Eine Internetplattform für
das Sprachenlernen* für die *Sprichwort-Plattform* entwickelt wurden (vgl. SWP;
Steyer 2012; Steyer und Ďurčo 2013; Ďurčo, Steyer und Hein 2017; Jesenšek 2011,
2013), und andererseits auf der im Folgenden vorgestellten korpusinformierten
Untersuchung.

    Der vorliegende Artikel bemüht sich um eine transparente Darstellung
der korpusinformierten Methodik für die Untersuchung des Slogangebrauchs
(Abschn. 3 und 4), des lexikografischen Prozesses, der Konzeption der Mikro-
struktur der Slogan-Artikel (Abschn. 5) sowie die verschiedenen Zugriffsmög-
lichkeiten in *OWID* (Abschn. 6). Darüber hinaus wird ein Einblick in die bis-
herigen lexikografischen Versuche der Sloganbeschreibung gegeben (Abschn. 2).

2.   Werbeslogans in Nachschlagewerken

Lexikografisch wurden Slogans bis dato zwar einerseits in klassischen oder on-
line-zugänglichen Zitatensammlungen (Büchmann 1864, ⁴²2001; Duden 12 (³2008);
Jeromin 1969; *Redensarten-Index*; *Liste geflügelter Worte*) und andererseits in Slogan-
lexika (Hars 2002) erfasst, allerdings wurden in diesen Nachschlagewerken
Slogans fast nie eigenständig und oft nicht systematisch sowie korpusbasiert
beschrieben. Interessanterweise waren in der Ausgabe des Dudenbands 12
Zitate und Aussprüche aus dem Jahr 2002 15 Werbeslogans, in der kommenden
Ausgabe aus dem Jahr 2008 bereits 31 Werbeslogans verzeichnet, was auf das
Interesse bzw. Offenheit der Sprachgemeinschaft für moderne Sprüche sowie
„auf die Präsenz und Tendenz zur Usualisierung von Slogans im aktuellen
Sprachgebrauch hinweist" (Polajnar 2019: 45); in den Wörterbuchartikeln des
Dudenbands 12 sind Werbeslogans nicht systematisch beschrieben und deren
Gebrauch ist nur relativ selten mit Belegen aus elektronischen Korpora veran-
schaulicht. Auch in den lexikografischen Online-Ressourcen sind nur verein-
zelte Werbeslogans verzeichnet. Zudem wird in den Wörterbuchartikeln dieser
Ressourcen eher auf die Entstehungsgeschichte als auf den Gebrauch im Alltag
bzw. auf den Korpusgebrauch eingegangen. Auch das umfangreichste Nach-
schlagewerk zu Werbeslogans von Wolfgang Hars (2002), *Das Lexikon der
Werbesprüche. 500 bekannte deutsche Werbeslogans und ihre Geschichte*, fokussiert
die werbegeschichtlichen Informationen und es wird nur punktuell auf ihren
Gebrauch eingegangen. Es kann also festgehalten werden, „dass bis dato
korpusinformierte Untersuchungen zu Slogans im aktuellen Sprachgebrauch
fehl[t]en" (Polajnar 2019: 14) und eine nach Kriterien der wissenschaftlichen
Lexikografie systematisch erarbeitete Onlinedokumentation von Slogans wie
die vorliegende als Desideratum galt. Dies hängt damit zusammen, dass
Werbeslogans bis dato nicht korpuslinguistisch systematisch untersucht
wurden, was die Grundlage für eine systematische, korpusinformierte, lexiko-
grafische Beschreibung und Dokumentation darstellt.[4]

    Werbeslogans wurden bis dato als Gegenstand unterschiedlicher For-
schungsdisziplinen behandelt, vorwiegend in der Werbesprachenforschung

und zwar als Bausteine der Werbung im Hinblick auf ihre Form, Inhalt und Funktion. Darüber hinaus wurden Werbeslogans als eigenständige Bausteine bzw. Texte im Hinblick auf ihre Einbindung in außerwerbliche Kontexte aus Sicht der Textlinguistik (vgl. Fix 1997, 2007; Janich 1997, 2019), Jugendsprachenforschung (vgl. Androutsopoulos 1997; Schlobinski 1989), Medienlinguistik (vgl. Betz 2006) sowie Phraseologie (vgl. Burger, Buhofer und Sialm 1982; Lüger 1999) untersucht (vgl. Polajnar 2019: 19-51). Hierbei entstanden einige textlinguistische und phraseologische Definitions- und Klassifikationsversuche, die den Slogan entweder als Werbespruch (Hemmi 1994: 62), Spruchtextsorte (Fix 2007: 464), geflügeltes Wort (Janich 1997) oder satzwertigen Phraseologismus (Lüger 1999; Burger 2015) auffassten; das Prinzip ihrer Wiederverwendung bzw. Verselbstständigung wurde im Hinblick auf Intertextualität (vgl. Fix 1997; Janich 1997) oder Phraseologisierung mit den Stufen Zitat, geflügeltes Wort, satzwertiger Phraseologismus (vgl. Burger, Buhofer und Sialm 1982: 56) perspektiviert. Die einzige korpusinformierte Untersuchung zum aktuellen Slogangebrauch außerhalb Werbung ist die Untersuchung von Polajnar (2019), die dem Modul *Werbeslogans* zugrunde liegt.

3.   Korpusinformierter Zugang zum Slogangebrauch außerhalb der Domäne Werbung

Im Mittelpunkt der Untersuchung, die dem Modul *Werbeslogans* zugrunde liegt, standen „ausgewählte deutschsprachige Werbeslogans mit hohem Wiedererkennungswert und einer Tendenz zur Usualisierung im aktuellen Sprachgebrauch" (Polajnar 2019: 11). Ihre lexikalische Verfestigung konnte korpusinformiert[5] anhand umfangreicher elektronischer Korpora (*DeReKo*) validiert und rekonstruiert werden. In die Untersuchung wurden 44 kodifizierte Werbeslogans miteinbezogen, die zwei Slogansammlungen mit Referenzcharakter entstammen (Duden 12 2008 und eine Liste der bekanntesten (Firmen-)Slogans vom Internetportal *slogans.de*). „Für die Beschreibung ihrer Verwendungsspezifik als eigenständige satzwertige Wortschatzeinheiten außerhalb der Domäne Werbung wird das Modell der usuellen Wortverbindungen sowie die korpuslinguistische Methodologie von Steyer (u. a. 2013, 2018) auf die Spruchtextsorte ,Slogan' angewandt" (Polajnar 2019: 11)[6] und in einem nächsten Schritt mit weiteren qualitativen und quantitativen Methoden (vgl. Keibel 2008; Polajnar 2012) verknüpft. Die Slogan-Muster und Slogan-Varianten wurden mit dem Analysewerkzeug zur Systematisierung von KWIC-Zeilen *lexpan* (2017) erarbeitet.

3.1   Iterative Phrasensuche mit engen und weiten Suchanfragen

Die Korpusvalidierung ausgewählter Slogans und deren Varianten sowie die Identifizierung von Slogan-Mustern erfolgt mit einer unterschiedlich elaborierten iterativen Phrasensuche im *Deutschen Referenzkorpus (DeReKo)*. Slogans wurden zunächst mithilfe einer komplexen Suchprozedur[7] via COSMAS II eigens analysiert und dann anhand der Korpusevidenzen qualitativ bewertet.

Als Beispiel seien die enge und die weite Suchanfrage[8] zum Slogan *Da weiß man, was man hat.* dargestellt und diskutiert. Die enge Suchanfrage dient der genauen Ermittlung von Korpusbelegen für die Kernform, z. B. *$da /+w1:1 weiß /+w1:1 man /+w1:1 was /+w1:1 man /+w1:1 hat*. In der Regel umfassen enge Suchanfragen die im Stichwort angeführten Wortformen des Slogans und einen geringen Wortabstand; sie orientieren sich an der Form des Stichworts. Man bekommt bei dieser Suche dann alle Belege mit genau dieser Kernform. Hierbei können die ermittelten Kernformen noch sehr eng an dem werblichen Ursprungs-kontext gebraucht werden, oder aber haben bereits eine allgemeinere Bedeu-tung entwickelt, wie die folgenden Belege aus dem Korpus zeigen.

(1)   HAZ09 die Vertrautheit des guten, alten „Tatorts". **Da weiß man, was man hat,**
        HMP12 Eiche, **da weiß man was man hat.** Ein echter Wert und "unkaputtbar".
        T12 2013 wird wieder Merkel gewählt **(da weiß man, was man hat)** und wir rutschen
        RHP13 Ich koche jedes Jahr Marmelade ein. **Da weiß man, was man hat,** und sie schmeckt
        NGACB Dann mache die Updates doch alle manuell. **Da weiß man, was man hat.**

Bei einer weiten Suchanfrage hingegen erweitert man den Wortabstand zwischen den Formen des Stichworts und/oder bezieht das ganze Flexions-paradigma ein. Mittels einer weiten Suchanfrage können lexikalische, aber auch syntaktische Varianten sowie Musterrealisierungen und dadurch Slogan-Muster erarbeitet werden; letztere geben des Weiteren die Auskunft über den invarianten Kern, lexikalische Füller sowie die Einschübstellen: *$da /+w1:1 (&wissen oder weiß) /+s0 was /+s0 &haben* (siehe 5i. Formvarianten und 5j. Ersetzung von Komponenten). Um eine passende weite Suchanfrage bei einem jeden Slogan zu fixieren, muss diese durch ein iteratives Verfahren feingetunt werden. So gibt die Auswahl der Konkordanzen der genannten weiten Such-anfrage Auskunft darüber, dass im Slogan *Da weiß man, was man hat.* aus-schließlich die beiden Indefinitpronomen *man* durch verschiedene Personal-pronomen ersetzt werden können, was bereits Indizien für den invarianten Kern *Da [wissen] X, was X [haben].* liefert.

(2)   U02 „Kinder, kauft Staatspapiere, **da wisst** ihr, **was** ihr **habt**
        BRZ06 pflanze ich lieber selber Bohnen, Erbsen und Möhren an. **Da weiß** man doch, **was** man **hat".**
        NGLD **Da weißt** Du, **was** Du **hast** — und musst nichts wegwerfen.
        T91 auf den Kleinstadtgeschmack abgestimmt. **Da weiß** der Alte, **was** er **hat.**

Aus methodischer Sicht können also bereits KWIC- und Volltext-Analysen wich-tige Hinweise über den invarianten Kern bzw. die Binnenstruktur eines lexika-lisch geprägten Slogan-Musters, die typischen Slogan-Varianten sowie Gebrauchs-restriktionen liefern. Besonders ertragreich scheint in solchen Fällen die iterative Suchheuristik (z. B. Ausschließen von Teilkomponenten u. a.) sowie KWIC-Ana-lysen und KWIC-Systematisierung mit Analyseprogramm *lexpan* (2017).

3.2   Exploration syntagmatischer Slogan-Muster mit *lexpan*

Mit dem Analysewerkzeug zur Systematisierung von KWIC-Zeilen *lexpan*[9] (2017)

erfolgte die systematische Analyse von Slogan-Varianten und syntagmatischen Slogan-Mustern und zwar mithilfe von exportierten KWIC- und Kookkurrenz-listen aus *DEREKO*.[10] Das einzelsprachenunabhängige Analyseprogramm dient „der explorativen Untersuchung von Festigkeit, Varianz, Slotbesetzungen und kontextuellen Einbettungsmustern syntagmatischer Strukturen" (ebd.).

Mithilfe von *lexpan* konnten die zunächst manuell erarbeiteten Ergebnisse in folgenden Bereichen automatisch verifiziert werden. Zum einen lassen sich mit dem Analysewerkzeug Slogan-Muster explorieren, indem Leerstellen ent-deckt und systematisch untersuchen werden. Zum anderen kann man auf diese Weise den Kontext von Slogans bzw. ihre kontextuelle Einbettung analysieren, indem man typische kontextuelle metakommunikative Marker und minimale lexikalische Einschübe systematisch untersucht. „Der Mehrwert von *lexpan* gegenüber manuellen Analysen erweist sich nicht nur in einer viel einfacheren, systematischen Identifikation von einbettenden Elementen, sondern auch von kontextuellen Einbettungsmustern." (Polajnar 2019: 65)

Um mit *lexpan* den Ikea-Werbeslogans *Wohnst du noch, oder lebst du schon?* auf Musterhaftigkeit bottom up anhand lexikalischer Musterrealisierungen zu rekonstruieren, muss man mit einer weiten Suchanfrage im virtuellen Zeitungs-korpus *noch /+w1:1 „oder" /+w3:3 schon* eine KWIC-Liste ermittelt. Nach dem Exportieren in *lexpan* wird die KWIC-Liste mit dem Suchmuster *# # noch oder # # schon* durchsucht. Die automatische Auswertung der Y-Leerstellen ergibt, dass diese auf Personalpronomen restringiert sind, darunter überproportional häufig auf das Personalpronomen *du-du* bzw. *Du-Du* (78,40 %). Daraus ergibt sich, dass der lexikalisch geprägte Slogan-Muster *X du noch, oder Y du schon?* eine prototypische Teilrealisierung des abstrakteren Musters darstellt. Allerdings weisen auch weitere Personalpronomina gewisse Vorkommenshäufigkeit auf: *Sie-Sie* (6,64 %), *ihr-ihr* (2,66 %), *sie-sie* (1,66 %), *wir-wir* (1 %), *er-er* (0,66 %) und *es-es* (0,66 %) (Abb. 2).

Abb. 2:   Schnappschuss der automatisch erstellten Lückenfüllertabelle für (0–3 Leerstellen) *noch oder* (0–3 Leerstellen) *schon* (Quelle: *lexpan*: Ausschnitt)

Analysiert man zugleich alle vier Slots des abstrakten Musters, ergibt die Füllerliste aus *lexpan*, dass der Slogan *Wohnst du/Du noch, oder lebst du/Du schon?* mit 15,28 % als der prototypische Vertreter gilt, da diese Verb-Pronomen-Kombination im Vergleich zu den anderen Kombinationen eine relativ hohe Vorkommenshäufigkeit aufweist (Abb. 3). Unter den häufigeren lässt sich auch die umgekehrte Kombination beider Verben *wohnen* und *leben* konstatieren, die als eine syntaktische Slogan-Variante aufzufassen ist: *Lebst du/Du noch, oder wohnst du/Du schon?* (2,66 %).

Abb. 3:   Schnappschuss der automatisch erstellten Lückenfüllertabelle für (0–3 Leerstellen) (0–3 Leerstellen) *noch oder* (0–3 Leerstellen) (0–3 Leerstellen) *schon* (Quelle: *lexpan*: Ausschnitt)

4.   *OWID*-Sloganliste

Vor der lexikografischen Beschreibung muss man sich zunächst den Selektionsmethoden bzw. der Sloganauswahl widmen. Die hiesige Stichwort-Liste der Werbeslogans stellt eine Auswahl der häufigsten Werbeslogans dar, die den Gegenstand der oben erwähnten korpusinformierten Untersuchung zum Slogangebrauch in Zeitungskorpora des *DEREKO* bildeten. Die ersten 30 Slogans mit höchster Vorkommenshäufigkeit wurden im nächsten Analyseschritt auf ihre „Satzwertigkeit" („satzwertige Phraseologismen" Lüger 1998, 1999) hin genauer untersucht, weil sich Satzwertigkeit bereits bei der Sprichwortidentifizierung als empirisch gut operationalisierbares Kriterium erwiesen hat (vgl. Steyer und Ďurčo 2013). Der nichtsatzwertige Gebrauch bzw. die Einbettung des Werbeslogans in die Satzstruktur hat nämlich oft seine Dekomposition zur Folge. Werden Werbeslogans, wenn auch im Korpus häufig vorkommend, also vorwiegend in die Satzstruktur eingebunden gebraucht, so wurden sie in das Modul nicht aufgenommen. Als Beispiel seien Belege zweier ausgesonderten Slogans *Das einzig Wahre.* (Warsteiner, 1972) und *Der Duft der großen weiten Welt!* (Peter Stuyvesant, 1959) genannt, die den nichtsatzwertigen Korpusgebrauch veranschaulichen.

(3)   NUN90 das ich liebe, das habe ich bis zur Langweile wiederholt, **das einzig wahre** Rimini ist das, was ich im Studio gebaut habe."
NUN90 Appell auch all jene erreicht, die glauben, unser System sei **das einzig wahre** auf dieser Welt.
T90 idealistischen Logik erscheint das fortschreitende Ganze als **das einzig Wahre,** welches von den "Achtundsechzigern" zwanzig Jahre lang

(4)   NUN90 "Vielleicht fehlt ein bißchen **der Duft der großen weiten Welt** im Vergleich zu anderen Berufen?"
T90 Die kannten **den Duft der großen weiten Welt** schon und rauchten den ganzen Abend Joints
T90 Die waren ja noch nie in Italien und Spanien und wollen **den Duft der großen weiten Welt** auch mal schnuppern.

Im Modul *Werbeslogans* sind nur solche satzwertigen Werbeslogans vorzufinden, die sich bereits von ihrem Ursprungskontext (Produktwerbung) entfernt haben und eine Tendenz zum „Weisheitssatz" aufweisen. Hierbei wurden in der *OWID*-Sloganliste (Abb. 4) Werbeslogans aus unterschiedlichen Jahrzehnten des 20. und 21. Jahrhunderts lexikografisch erfasst:

— aus den 20ern: *Nie war er so wertvoll wie heute.* (Klosterfrau Melissengeist, 1925),
— aus den 50ern: *Mach mal Pause*-Serie (Coca-Cola, 1955), *Mit fünf Mark sind Sie dabei.* (ARD-Fernsehlotterie, 1956), *Er läuft und läuft und läuft …* (Volkswagen, 1959) und *Der Duft der großen weiten Welt!* (Peter Stuyvesant, 1959),
— aus den 60ern: *Wer wird denn gleich an die Luft gehen?* (HB, 1960), *Alle reden vom Wetter, wir nicht.* (Deutsche Bahn, 1966), *Da weiß man, was man hat.* (Volkswagen, 1969),
— aus den 70ern: *Quadratisch, praktisch, gut.* (Ritter Sport, 1970), *Die zarteste Versuchung, seit es Schokolade gibt.* (Milka, 1971), *Das einzig Wahre.* (Warsteiner, 1972), *Es gibt viel zu tun. Packen wir's an.* (Esso, 1974), *Nicht immer, aber immer öfter.* (Clausthaler, 1979),
— aus den 80ern: *Man gönnt sich ja sonst nichts.* (Malteserkreuz Aquavit, 1985), *Wir machen den Weg frei.* (Volks- und Reiffeisenbanken, 1988),
— aus den 90ern: *Nichts ist unmöglich.* (Toyota, 1992), *Ich bin doch nicht blöd.* (Media Markt, 1996) oder *Ja is' denn heut' scho' Weihnachten.* (E-Plus, 1998) und
— aus dem 21. Jh.: *Geiz ist geil!* (Saturn, 2001), *Wohnst du noch oder lebst du schon*? (Ikea, 2002).

Die empirischen Untersuchungen haben gezeigt, dass Werbeslogans oft, aber nicht immer eine auffällige sprachliche Struktur aufweisen wie *Wohnst du noch oder lebst du schon*? (Ikea, 2002). In einigen Fällen basieren sie allerdings auf völlig regulären, strukturell unauffälligen Sätzen (z. B. *Ich liebe es* (McDonald's, 2003), *Mit fünf Mark sind Sie dabei.* (ARD-Fernsehlotterie, 1956)), die zum Teil umgangssprachlichen oder dialektalen Charakter aufweisen (z. B. *Ich bin doch nicht blöd.* (Media Markt, 1996), *Es gibt viel zu tun. Packen wir's an.* (Esso, 1974) sowie *Ja is' denn heut' scho' Weihnachten* (E-Plus, 1998)). Erst durch die Verknüpfung mit einer Marke bzw. einem Produkt und durch vielfaches Wiederholen treten sie

aus dem Fluss der Werbekommunikation heraus. Die Sprachgemeinschaft ent-
scheidet dann schließlich darüber, ob der Slogan genug „Spruch-Potenzial"
hat, d. h. Alltagssituationen, Verhaltensweisen und Normen plastisch kom-
mentiert und auf den Punkt bringt und damit die Chance besitzt, in den
Sprachbestand auf Dauer überzugehen.
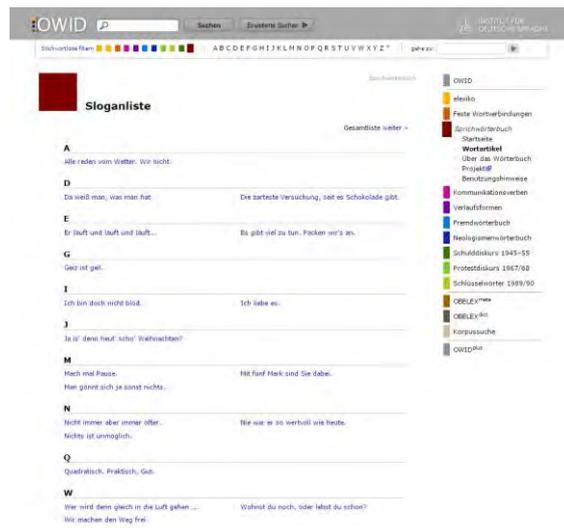
Abb. 4:   *OWID*-Sloganliste

5.   Aufbau der Slogan-Artikel

Die korpusbasierte Beschreibung umfasst folgende Bausteine: „Kernform",
„Basiskomponenten", „Äquivalente in anderen Sprachen" (falls vorhanden),
„Suchanfragen für Recherche im Korpus", „Geschichte", „Bedeutung", „Ge-
brauchsbesonderheiten", „Formvarianten", „Ersetzung von Komponenten",
„Typische Verwendung im Text" und „Vorkommen in Nachschlagewerken".
Die Slogan-Artikel unterscheiden sich von den anderen Sprichwörterbuch-
Artikeln durch folgende Strukturmerkmale: In den Bausteinen „Kernform",
„Äquivalente" und „Geschichte" (siehe dazu 5a, 5c und 5e) wird auf die Ent-
stehungsgeschichte als (internationaler) Firmen-, Marken- oder Produktslogan
verwiesen; zudem wird am Ende der Slogan-Artikel im Baustein „Vorkommen
in Nachschlagewerken" (5k) ihre bisherige lexikografische Erfassung eviden-
tiert. Im Folgenden sollen die einzelnen Bausteine der Slogan-Artikel anhand
von Beispielen dargestellt und kommentiert werden.

a. Kernform

Obwohl man bei Werbeslogans intuitiv in die Versuchung geraten könnte, den

Originalslogan als Kernform anzusetzen, zeigt die empirisch fundierte und hiermit methodisch sichere Korpusanalyse (Abschn. 3) ein anderes Bild. Die Kernform stellt folglich die im Korpus häufigste satzwertige Form des Werbeslogans dar; diese ist im Vergleich zum Originalslogan oft ohne Angabe des Markennamens. Die Korpusanalyse, mit welcher die auffällig rekurrente Vorkommensform identifiziert wird, hat gezeigt, dass Markennamen oft eine Hürde bei der Verselbstständigung von Slogans darstellen. Dies ist auf ihre Identifikationsfunktion des Unternehmens bzw. der Marke und hiermit mit dem Verweis auf die Ursprungsdomäne zurückzuführen. So können Werbeslogans, bei denen der Markenname Teil der Satzstruktur ist, erst auf der Musterebene eine usuelle, situationsunabhängig-abstrakte (nicht-werbliche) Bedeutung erlangen, wie dies beim Haribo-Slogan *X macht Kinder froh.* deutlich wird.

Bei den meisten Werbeslogans entspricht die Kernform dem Slogan ohne den Markennamen. Zudem liegen Slogan-Artikel mit Kernformen vor, die eine verkürzte Form des Originalslogans darstellen: Z. B. wurde die Kernform vom Werbeslogan für die Deutsche Bahn aus dem Jahr 1966 *Alle reden vom Wetter. Wir nicht. Fahr lieber mit der Bundesbahn.* anhand von Korpusvorkommen auf *Alle reden vom Wetter. Wir nicht.* angesetzt. Unter der Kernform wird im Werbeslogan-Artikel deshalb immer auch der Originalslogan mit dem Markennamen aufgeführt (Abb. 5).

Abb. 5:   OWID-Sloganartikel: Stichwort *Alle reden vom Wetter. Wir nicht.*

Wie die Sprichwörter weisen auch die meisten Werbeslogans eine eindeutige Kernform auf. Bei einigen Werbeslogans liegen jedoch mehrere Varianten vor, bei denen sich eine als prototypisch erweist (Kernform) und die anderen unter Formvarianten aufgeführt werden (siehe dazu auch h) (vgl. Steyer und Ďurčo 2013).

Kernform: *Es gibt viel zu tun. Packen wir's an.*

Formvarianten: *Es gibt viel zu tun. Packen wir es an.*; *Es gibt viel zu tun. Packt es an.*; *Packen wir es an.*; *Packen wir's an.*; *Packen Sie es an.*

Bei Werbeslogans, die beispielsweise verselbstständigte Sätze aus Werbespots

darstellen (gebundene Slogans) und nur gesprochen realisiert werden, konkurrieren mehrere (orthografische) Varianten miteinander, von denen keine als prototypisch gelten kann. In einem solchen Fall muss eine kompetenzbasierte Entscheidung getroffen werden. Als Beispiel sei der Werbeslogan für E-Plus genannt:

Kernform: *Ja is' denn heut' scho' Weihnachten?*

Formvarianten: *Ja is' denn heut' schon Weihnachten?*; *Ja ist denn heut' scho' Weihnachten?*; *Ja ist denn heute schon Weihnachten?*

b. Basiskomponenten

Alle lexikalischen Komponenten des Slogans (außer bestimmtem und unbestimmtem Artikel) werden zwecks Verlinkung mit dem *elexiko-Wörterbuch* in *OWID* und der Ermöglichung unterschiedlicher Zugriffsmöglichkeiten gesondert ausgezeichnet.

c. Äquivalente in anderen Sprachen

Diese Rubrik wird in all jenen Fällen bearbeitet, in denen Webseiten unterschiedlichen Charakters mit Äquivalenten in anderen Sprachen gefunden wurden (Wikipedia, Firmenseiten usw.): In der Regel handelt es sich um Werbeslogans international agierender Unternehmen, wie z. B. Milka: *Die zarteste Versuchung, seit es Schokolade gibt.*, McDonald's: *Ich liebe es.* oder Media Markt: *Ich bin doch nicht blöd.*, die wegen europaweiten oder weltweiten Präsenz Marketing bedingte Äquivalente in unterschiedlichen Sprachen aufweisen. Bei den Äquivalenten internationaler Slogans handelt es sich meist um Teiläquivalente, vereinzelt wurden auch semantische Äquivalente und Nulläquivalente konstatiert (Polajnar 2016). Ob diese in den jeweiligen Sprachen genauso wie der Originalslogan in außerwerblichen Kontexten Verwendung finden, müsste einzeln für jede Äquivalente wie im Deutschen korpusbasiert überprüft werden. Eine exemplarische Untersuchung von Sloganäquivalenten (Polajnar 2016) hat jedoch zeigen können, dass viele populäre internationale Slogans nicht nur in der Ausgangssprache bzw. im Deutschen, sondern auch in den anderen Sprachen erneut in außerwerblichen Kontexten Verwendung finden und sogar Variationsmuster bilden (z. B. für Snickers im Slowenischen: *X si ful drugačen.* ('X bist du ganz anders') sowie für Red Bull im Englischen: *.X gives you wings.* usw.).

d. Suchanfrage für Recherche im Korpus

Diese bereits im großen *OWID*-Sprichwörterbuch integrierte Angabe soll der schnellen Auffindbarkeit von authentischen Sprachbelegen im Deutschen Referenzkorpus (*DeReKo*) dienen. Die Suche nach komplexen Sätzen stellt gerade für einen unerfahrenen Nutzer oft eine erhebliche Hürde dar, muss er sich doch einer manchmal recht komplizierten Syntax bedienen. Die hinterlegten

Suchanfragen können nun ganz problemlos in das COSMAS II-Suchfenster kopiert werden. Ein neues *OWID*-Feature wird demnächst diese Suche noch weiter erleichtern. Der Nutzer muss dann nur noch auf die Suchanfrage klicken, und die Korpusanalyse startet automatisch im Hintergrund. Der Nutzer kann so alle Originaltextstellen, in denen der Slogan vorkommt, immer auf dem jeweils aktuellen Stand von DEREKO erhalten. Es werden eine enge und eine weite Suchanfrage für weitere eigenständige Recherchen im Korpus aufgeführt (vgl. Abschn. 3.1).

e. Geschichte

Anders als bei den Sprichwörtern war es bei den Slogans wichtig, auf ihren ursprünglichen Gebrauch in der Werbung bzw. ihre Entstehungsgeschichte sowie auf ihre domänenspezifische Bedeutung zu verweisen. Weil der Ursprung bzw. die Quelle von Werbeslogans nachgewiesen werden kann, werden Slogans häufig als zeitspezifische geflügelte Worte der jüngsten Vergangenheit bezeichnet (vgl. Janich 2010: 61). Doch die Slogan-Artikel zeigen, dass Slogans nicht nur zitiert werden, sondern Varianten und vor allem Muster mit Leerstellen aufweisen, die einem ermöglichen, sie durch das Aufgreifen von Schlüsselwörtern an beliebige Kontexte anzupassen. Dies verweist eher auf ihren ausgeprägten musterhaften Gebrauch (vgl. Bubenhofer 2009).

In diesem Teil des Slogan-Artikels werden also Informationen zur Entstehung des Werbeslogans aus Sekundärquellen gegeben (Nachschlagewerke, Webportale und einschlägige Webseiten usw.). Angereichert werden diese Texte durch DEREKO-Belege, in denen die Entstehung oder der Ursprungskontext des jeweiligen Slogans explizit thematisiert wird und Werbeslogans in Zitatform gebraucht werden. Dadurch unterscheidet sich auch der Korpusgebrauch von Slogans in diesen Belegen vom Korpusgebrauch in allen weiteren Belegen, da hier der Werbeslogan noch keine usuelle, situationsunabhängig-abstrakte (nicht-werbliche) Kernbedeutung aufweist, wie sie in weiteren Teilen der Slogan-Artikel genau beschrieben wird (siehe f). Zudem liegen hier Links auf Youtube-Videos mit den Originalwerbespots (wenn vorhanden) vor. Beim Werbeslogan der Firma Ritter Sport *Quadratisch, praktisch, gut.* werden in dieser Rubrik eine Vorgeschichte samt Werbespot und ein Beleg, der auf die Werbung rekurriert (Abb. 6), dargestellt:

Geschichte

Die Firma Ritter Sport wirbt noch heute mit diesem Werbeslogan für die Schokolade der Marke Ritter Sport, die für ihre quadratische Form bekannt ist. "Es gibt die Schokolade in dieser Form [...] bereits seit 1932, der populäre Werbespruch wurde aber erst im Jahr 1970 entwickelt. Heute zitiert man ihn im Zusammenhang mit den verschiedensten Produkten oder Ideen meist dann, wenn man deren praktische Seite lobend hervorheben möchte" (Duden 12 2008: 434).

Werbespot 1994 auf Youtube

Korpusbeleg

Zu bundesweiter Bekanntheit brachte es Ritter Sport ab 1970 mit der ersten Joghurtschokolade und den drei Worten: quadratisch, praktisch, gut. Jede Sorte erhielt ihre eigene Farbe und eine Verpackung, die sich durch einfaches Knicken öffnen lässt. (Die Rheinpfalz, 22.02.2012, S. 7; Eine süße Idee wird 100 Jahre alt)

Abb. 6:   *OWID*-Sloganartikel: Geschichte *Quadratisch, praktisch, gut.* mit dem Link auf den Werbespot

f. Bedeutung

Die Bedeutungserklärungen basieren durchweg auf den authentischen Korpusvorkommen. Hier geht es um die usuelle, situationsunabhängig-abstrakte (nicht-werbliche) Kernbedeutung, die in allen Korpusverwendungen nachweisbar sein muss. Unter „Bedeutung" wird der verallgemeinerte semantische Kern erfasst, der für alle üblichen Vorkommen dieses Slogans gleichermaßen zutrifft (vgl. *SWB*). Hierbei verwenden wir eine weitgehend standardisierte Beschreibungssprache, die bereits im EU-Projekt Sprichwort (vgl. Steyer 2012 und Sprich-Wort-Projekt) aus fremdsprachendidaktischen Gründen eingeführt wurde, beispielsweise einleitende Formulierungen wie „Sagt man dafür, dass ..."; „Sagt man, wenn ...". Die Bedeutungsbeschreibung soll anhand zweier Beispiele illustriert werden. Für den Slogan *Da weiß man, was man hat.* lautet die Bedeutung wie folgt: „Sagt man, wenn man lieber auf Bewährtes setzt, als neue, aber möglicherweise riskantere Dinge anzustreben." (SWB-SL) Dazu werden Belege aus dem Korpus aufgeführt.

(5)   *Wie Gertrud und Reiner Baum machen es viele der inzwischen mehr als 20 000 deutschen Hausbesitzer an Floridas Golfküste: Außerhalb des eigenen Urlaubs vermieten sie ihre Ferien-Immobilie wochenweise — am liebsten an Landsleute: "**Da weiß man, was man hat!**"* (Rhein-Zeitung, 28.02.1997; florida)

Der Slogan *Quadratisch, praktisch, gut* wird folglich paraphrasiert: „Sagt man, wenn etwas ohne großen Aufwand seine Funktion erfüllt und Qualität hat." (SWB-SL)

(6)   *Auf- und Umbau funktionieren nach dem Baukastenprinzip: Die dünnen Holzplatten werden mithilfe einer neuartigen Kunststofffolie zum Würfel verspannt — fertig ist der Beistell- oder Nachttisch. Merkmal: **quadratisch, praktisch, gut,** ganz ohne Schrauben, Bohren oder Hämmern.* (Hannoversche Allgemeine, 29.05.2010; Mobil mit Modulen)

Bedeutungsbeschreibung fällt dann anspruchsvoller aus, wenn ein Slogan überwiegend in den Kontexten vorkommt, die noch stark mit der Domäne Werbung verknüpft sind. In einigen Fällen konstituiert sich die abstrakte Bedeutung erst auf der Musterebene. So ist die Verwendung des Deutsche-Bahn-Slogans *Alle reden vom Wetter. Wir nicht.* stark an den Wetter-Kontext gebunden. Erst durch die Ersetzung des Lexems *Wetter* durch andere Nomina

wie *Rezession* oder *Benzinpreise* lässt sich eine generelle Bedeutung festschreiben: „Sagt man dafür, dass man anders handelt als die Mehrheit und damit vom Gängigen abweicht." (SWB-SL)

(7)   *Heulen und Zähneklappern wie in vielen anderen Wirtschaftsbereichen ist in der Spielwarenbranche derzeit nicht angesagt:* **"Alle reden von Rezession — wir nicht"**, *betont DSVI-Geschäftsführerin Printzen mit Verweis auf insgesamt leichte Umsatzzuwächse.* "Allen Unkenrufen zum Trotz konnte sich der Spielwarenmarkt in der ersten Jahreshälfte gut behaupten", *unterstreicht auch Marktforscherin Gabriele Eberl.* (Nürnberger Nachrichten, 26.10.2001, S. 11; Spielwarenbranche hofft auf Umsatzplus im Gesamtjahr — Warten auf Weihnachten)

Ein vergleichbares Phänomen lässt sich beim Ikea-Slogan *Wohnst du noch oder lebst du schon?* feststellen. Der Slogan wird noch recht häufig in Verbindung mit dem Wohnen und der Philosophie des Möbelkonzerns Ikea gebraucht. Auch hier ist die abstraktere Bedeutung erst durch andere Verbfüller gegeben: *Frierst du noch oder heizt du schon? Isst du noch oder genießt du schon?* Die Bedeutung kann dan wie folgt festgeschrieben werden: „Sagt man dafür, dass zu fragen ist, ob ein Zustand noch erstrebenswert oder ein anderer zu bevorzugen ist." (SWB-SL)

(8)   *Nein, noch geht es nicht um die Weihnachtsgeschenke — auch wenn im Handel längst wieder Lebkuchen und Spekulatius Einzug gehalten haben und familienintern mitunter schon diskutiert wird, um wessen Weihnachtsbaum man sich an Heiligabend versammeln wird. Vorerst wird allenthalben die Frage diskutiert:* **„Frierst du noch oder heizt du schon?!"** *Angesichts der Wetterkapriolen, die sich vom Frühling über den Sommer nahtlos in den Herbst gerettet haben, hat man in den letzten Wochen kleidungstechnisch dem Zwiebellook gefrönt: morgens Wollpullover, mittags T-Shirt, abends Daunenjacke. Aber jetzt schon heizen?* (Mannheimer Morgen, 01.10.2010, S. 17)

g. Gebrauchsbesonderheiten

Unabhängig von der Grundbedeutung liegen bei Slogans zudem Gebrauchsbesonderheiten vor; diese repräsentieren das Typische in den Korpusbelegen und können nur in bestimmten konkreten Gebrauchssituationen auftreten. Unter „Gebrauchsbesonderheiten" werden konnotative, pragmatische und domänenspezifische Aspekte des Slogangebrauchs berücksichtigt, die in den Belegen oft zu beobachten sind, aber nicht für alle Vorkommen verallgemeinerbar sind: Die Beschreibungen der Gebrauchsbesonderheiten werden durch folgende Formulierungen eingeleitet: „(In den Korpusbelegen wird) häufig ...", „typischerweise" oder „in bestimmten Korpusbelegen" (*SWB*). Für den Slogan *Da weiß man, was man hat.* lautet eine der Gebrauchsbesonderheiten wie folgt: „In den Korpusbelegen wird mit dem Slogan häufig thematisiert, dass man landwirtschaftliche Produkte aus der eigenen Heimat oder aus vertrauter Herkunft bevorzugt." (SWB-SL)

(9)   *Es war 1989. Die Öffentlichkeit erfuhr erstmals vom Thema BSE, dem Rinderwahnsinn. Da konnte die schwangere Dorothe Lengert, obwohl keine Vegetarierin, "kein Fleisch mehr sehen", zumal auch eine Krankheit bei Schweinen grassierte. Die in Hochscheid geborene Försterstochter und ihr Mann Thomas, ein aus Hinzerath stammender Betriebsschlosser, suchten nach Alternativen. Denn Fleisch sollte auch weiterhin auf den Tisch. Da blieb nur die Haltung eigenen Viehs: "**Da weiß man, was man hat!"*** (Rhein-Zeitung, 15.02.1997; Altes Glanvieh kommt — durch BSE neu in Mode)

Der Slogan *Quadratisch, praktisch, gut.* weist ein ganzes Spektrum von Gebrauchs-besonderheiten auf (Abb. 7):

Abb. 7:   *OWID*-Sloganartikel: Gebrauchsbesonderheiten zu *Quadratisch, prak-tisch, gut*

h. Typische Formvarianten

Formvarianten (h) gehören zusammen mit Ersetzungen von Komponenten (i) zur Varianz, die auf Usualität basiert. Unter Formvarianten sind Informatio-nen zu allen usuellen Typen von formalen Systemvarianten einzelner Slogan-Komponenten oder des ganzen Slogans subsumiert (z. B. morphologische oder grammatische Varianten, orthographische Varianten). Es sind nur rekurrete Varianten verzeichnet, die bei relativ wenigen Slogans vorzufinden sind (vgl. *SWB*). Als Beispiel sind die morphologischen Formvarianten zum Coca-Cola-Slogan *Mach mal Pause.* sowie die lexikalischen Formvarianten zum Slogan *Mit fünf Mark sind Sie dabei.* zu nennen:

Slogan-Kernform: *Mach mal Pause.*
Morphologische Formvarianten: *Macht mal Pause.* und *Machen Sie mal Pause.*

Slogan-Kernform: *Mit fünf Mark sind Sie dabei.*
Lexikalische Formvarianten: *Mit fünf Euro sind Sie dabei.*

(10) *Machen Sie aus dem Möbeleinkauf ein kleines Erlebnis. Gehen Sie entspannt durch die einzelnen Abteilungen oder flanieren Sie von einem Einrichtungsgeschäft zum nächsten. Setzen Sie sich nicht selbst unter Druck.* **Machen Sie mal Pause** *und gehen zwischendurch etwas essen.* (Mannheimer Morgen, 22.10.1995; Wer nicht handelt, ist selber schuld)

Der Entstehung nach können Slogan-Varianten bei Werbeslogans unabhängig vom Variantentyp in zwei Gruppen unterteilt werden:

(a) domänenspezifische Slogan-Varianten, die im Rahmen der Werbekampagne(n) entstehen:
— Zu *Mit 5 <u>Mark</u> sind Sie dabei.* enstand wegen Währungsänderung eine lexikalische Variante *Mit 5 <u>Euro</u> sind Sie dabei.* (Fernsehlotterie)
— Zu *Haribo macht Kinder froh.* entstand wegen Zielgruppenerweiterung eine syntaktische Variante *Haribo macht Kinder froh <u>und Erwachsene ebenso</u>.* (Haribo)

(b) Slogan-Varianten, die im *DEREKO* identifiziert werden:
— Zum gesprochensprachlich markierten Slogan von E-Plus *Ja, is´ denn heut´ scho´ Weihnachten?* eine orthografische Variante *Ja, <u>ist</u> denn <u>heute schon</u> Weihnachten?*
— Zu *Nicht immer, aber immer öfter.* eine lexikalische Variante *Nicht immer, aber immer <u>öfters</u>.*
— Zu *Wer wird denn gleich in die Luft gehen?* syntaktisch-lexikalische Varianten *Aber halt — Wer wird denn gleich ...* sowie *Warum denn nicht gleich in die Luft gehen?*

i. *Typische Ersetzung von Komponenten*

Hier sind Variationsmuster verzeichnet, bei denen eine Leerstelle unterschiedlich besetzt wird. Die Ersetzung von Komponenten wird durch die Korpusanalyse ermittelt; hierbei muss das Kriterium der Verallgemeinerbarkeit erfüllt werden (vgl. Steyer 2013). Die Muster-Angaben enthalten feste lexikalische Komponenten (fester lexikalischer Kern) und so genannte Füllerangaben. Im Folgenden sind Beispiele dafür aufgeführt, wie die varianten Stellen im Korpus gefüllt werden (Abb. 8).
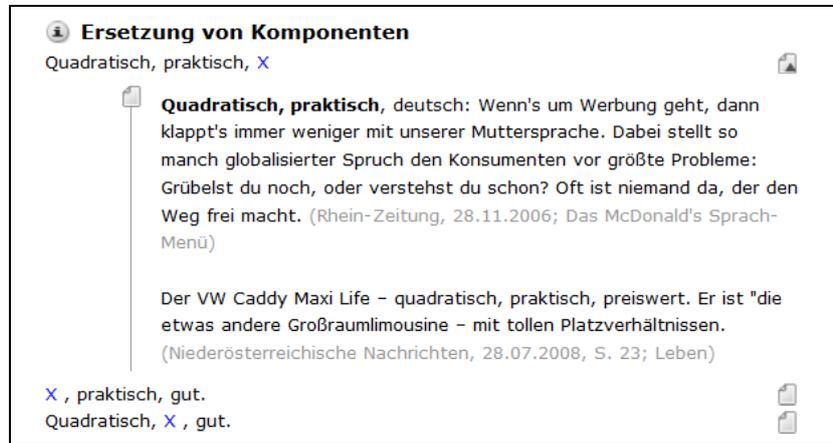
Abb. 8:   *OWID*-Sloganartikel, Ersetzung von Komponenten zu *Quadratisch, praktisch, gut*

Die lexikalischen Ersetzungen können auf unterschiedlichen Ebenen und mit unterschiedlichen Effekten erfolgen, z. B.:

(a) Ersetzungen unter Beibehaltung der Sloganbedeutung
  — durch formale Ersetzungen, z. B. *Da weiß man, was man hat.*
    *-> Da weiß **ich/sie/Frau**, was **ich/sie/Frau** habe/hat.*
  — durch andere Lexeme, z. B. *Er läuft und läuft und läuft ->* (*VW-Käfer/ Auto/AC Milano*) *läuft und läuft und läuft*.

(b) Ersetzungen, die unter Beibehaltung der Sloganbedeutung auf andere Sachverhalte referieren, z. B.
  — *Die zarteste Versuchung, seit es Schokolade gibt -> Die zarteste Versuchung seit es **Schinken/Männer/Parteien** gibt*;

(c) Ersetzungen, die die Sloganbedeutung verändern, z. B.:
  — *Alle reden vom Wetter, wir nicht -> Alle reden von/vom **Zukunft/Globalisierung/Superwahljahr**, wir **auch**; Alle reden von (der) **Krise/Fußball**, wir **nicht***.
  — *Es gibt viel zu tun. Packen wir es an -> Es gibt viel zu tun, **lassen wir es sein/warten wir es ab/fangt schon mal an***.
  — *Ja, is' denn heut schon Weihnachten -> Ja, is' denn heute schon **Rosenmomtag/1. April/Wahlkampf*** (Funktionswandel: Ausdruck von unverhoffter Freude -> Ausdruck von Verwunderung)

j.   *Typische Einbettungen in den Text*

Die Markierung von Slogans im neuen Kontext erfolgt durch Marker, Anführungszeichen, oft aber ohne jegliche Kennzeichnung. Diese Angabe erfasst

jedoch ausschließlich auffällige sprachliche Phänomene in der unmittelbaren sprachlichen Umgebung eines Slogans. Solche Einbettungen steuern häufig die Interpretation auf maßgebliche Weise. Beispiele für typische Einbettungen sind „fiktive Antworten" wie die folgende Negation:

(11)   *Neuschnee in der Lüneburger Heide!* **Ja, is denn heut scho Weihnachten?** <u>Nö</u>, *aber die erste Skihalle Norddeutschlands eröffnete gestern in Bispingen. (Hamburger Morgenpost, 21.10.2006, S. 17; Ein Auftakt zum Ausschütten)*

Zahlreiche Slogans werden im neuen Kotext nicht nur durch domänen- bzw. werbespezifische Termini (*Werbeslogan, Werbekampagne, Werbespruch* etc.) einge-bettet, sondern ähnlich wie Sprichwörter durch Marker wie *unter dem Motto, Devise* usw. Steyer kommt anhand der korpusbasierten Untersuchung von 2000 Sprichwörtern zum Schluss, dass es in der Sprachgemeinschaft „durchaus ein ausgeprägtes Sprecherbewusstsein von ‚Sätzen' gibt (nicht im grammatischen Sinne, sondern im Sinne einer funktional vollständigen Einheit)" (Steyer 2013: 348). Allerdings scheint die Klassifizierung dieser als *Sprichwort, Motto, Devise* usw. für die Sprecher/-innen unwichtig. In den Sloganartikeln werden die werbespezifischen oder allgemeinen Marker nur dann vermerkt, wenn diese selbst ein Muster bilden, wie beispielsweise die Vergleiche mit einem Substan-tiv *X à la* oder *X wie* bei Slogan *Geiz ist geil*.

(12)   *Und auch die anderen Fraktionen im Rathaus haben sich bei der Wahl des neuen Bürgermeisters am 1. Oktober **unter dem Motto** "Da weiß man, was man hat" mit großer Mehrheit für den, wie damals zu hören war, "sachkundigen, fairen und zu Kompromissen bereiten Kommunalpolitiker" ent-schieden. (Frankfurter Allgemeine, 04.11.1997; Der neue Bürgermeister ist jetzt "an Bord")*

(13)   *Die »MarktHallen« am Hauptgüterbahnhof stemmen sich gegen **Schnell-und-Billig-Trends à la** »Geiz ist geil«: Die Geschäftsleute wollen mit hochwertiger Qualität, die natürlich ihren Preis hat, und ihrem Hintergrundwissen die Kundschaft überzeugen. »Wir drücken niemandem etwas auf, son-dern wir geben Erklärungen zu unseren ausschließlich biologischen Produkten«, sagt »Markt-Hallen«-Betreiber und Käseverkäufer Jürgen Würth. (Nürnberger Nachrichten, 15.12.2003; Einkauf mit Genuss — "MarktHallen" wollen Dornröschenschlaf beenden)*

Als sprachlich auffällig werden unter typischen Einbettungen bei Slogans zudem markenspezifische Marker wie *Ikea-Parole, Lotterie-Weisheit, Toyota-Prinzip* genannt, also Bildungen mit Bindestrich-Komposita.

(14)   *Mit fünf Mark ist man bei der Verbandsgemeinde dabei*
*Man muss nicht immer mit der Zeit gehen — das meint offensichtlich die Verbandsgemeinde-Verwaltung Kirn-Land und hält es **mit der alten Lotterie-Weisheit "Mit fünf Mark sind Sie dabei".** Das sagt jedenfalls die Aufschrift an der Schrankenanlage vor dem Verwaltungsgebäude an der Bahnhofstraße aus. Doch keine Angst: Der Schlagbaum öffnet sich in diesen Euro-Zeiten auch für Nicht- Mark-Besitzer. Und wenn es keinen stört, muss man in Zeiten leerer kommunaler Kassen schließlich auch eine müde Mark für ein neues Schild ausgeben. (Rhein-Zeitung, 17.11.2004; Mit fünf Mark ist man bei der...)*

k. Vorkommen in Nachschlagewerken

Diese Angabe dient dokumentarischen Zwecken, indem vermerkt wird, ob ein Slogan bereits in irgendeiner Form kodifiziert ist.

Vorkommen in Nachschlagewerken

– Duden 12: Zitate und Aussprüche. 2008. Mannheim, Leipzig, Wien, Zürich.
– Hars, Wolfgang (2002): Nichts ist unmöglich! Lexikon der Werbesprüche. 500 bekannteste Werbeslogans und ihre Geschichte. München, Zürich.

Abb. 9:   *OWID*-Sloganartikel: Vorkommen in Nachschlagewerken

6.   Suche und Vernetzung innerhalb des *OWID*-Systems

Die Sloganartikel sind auf unterschiedliche Weise mit anderen Inhalten in *OWID* verbunden:

Bestimmte Slogans werden direkt mit jenen Sprichwortartikeln (*SWB*) verlinkt, mit denen sie in einem wie auch immer gearteten Verwendungszusammenhang stehen (vgl. Abb. 10):

– auf minimalistische, funktional-sachliche Architekur in Anlehnung an die Bauhaustradition verwiesen.

> Kommentar verbergen ×   Belege anzeigen »

In Bezug auf den Aspekt einer minimalistischen, funktional ausgerichteten Formensprache weist der Slogan eine Gebrauchsverwandschaft mit der Sprichwort *«Weniger ist mehr»* auf.

Abb. 10:   Screenshot des OWID-Sloganartikels *Quadratisch, praktisch, gut. -> Link zu Weniger ist mehr*

Andere Beispiele für Verbindungen zu Sprichwörtern aus dem *OWID-Sprichwörterbuch* (*SWB*) sind folgende:

***Nichts ist unmöglich*:**
   Wer wagt, gewinnt; Der Glaube versetzt Berge; Den Mutigen gehört die Welt; Sag niemals nie, Beharrlichkeit führt zum Ziel, Frisch gewagt ist halb gewonnen

Im Bedeutungsaspekt des „Erst-Mal-Abwartens"
Man sollte sich nicht zu freuen; Noch ist nicht aller Tage Abend; Man soll das Fell des Bären nicht erlegen, bevor er erlegt ist; Eine Schwalbe macht noch keinen Sommer

Nicht immer, aber immer öfter
Unverhofft kommt oft; Was lange währt, wird endlich gut; Steter Tropfen höhlt den Stein; Geduld bringt Rosen

Wer wird denn gleich in die Luft gehen …
In der Ruhe liegt die Kraft; Eile mit Weile

Durch die Auszeichnung der Basiskomponenten wird zum einen eine Verlinkung zu dem entsprechenden Einwort-Artikel in *elexiko* hergestellt, zum anderen kann die *OWID*-Suche auf die Basiskomponenten zurückgreifen. So erhält man beispielsweise bei der Komponente *gut* folgende Suchergebnisse (Abb. 11).

Abb. 11: Suche nach *gut* in *OWID* http://www.owid.de/suche/wort?wort=gut

Des Weiteren ist es möglich, alle Slogans und Sprichwörter nach einer Basiskomponente zusammenstellen zu lassen (Abb. 12):

Abb. 12:   Liste der Basiskomponenten (im Artikel Quadratisch auf *gut*)

OWID sieht schließlich die Möglichkeit vor, unterschiedliche Stichwortlisten anzeigen zu lassen. So kann man den Slogan *Da weiß man, was man hat* als Teil des Sprichwörterbuchs sehen. Es ist aber auch möglich, die gesamte *OWID*-Stichwortliste abzurufen.

6.   Schlussbemerkungen und Ausblick

Die dargestellte lexikographische Behandlung von Werbeslogans im Rahmen von OWID ist (in der germanistischen Lexikographie) ein Novum und trägt zur lexikographischen Behandlung von polylexikalen Lexikon-Einheiten bedeutend bei; vor allem aus der Sicht der neueren Betrachtungen des Lexikons aus der Perspektive der Konstruktionsgrammatik. Mit dem hier vorge-

stellten lexikografischen Prozess, der zum Modul *Werbeslogans* führte, verfolgten wir das Ziel, den Korpusgebrauch von aktuell gebräuchlichen deutschen Werbeslogans nach Kriterien der wissenschaftlichen Lexikografie aufzubereiten und als Onlinedokumentation nachhaltig zur Verfügung zu stellen. Da systematische korpusbasierte (Online-)Dokumentationen von Werbeslogans bis dato fehlen, dürfte das beschriebene Modul ein innovatives Konzept zur lexikografischen Erfassung und linguistischen Beschreibung der Werbeslogans darstellen. Anhand zahlreicher Korpusbelege konnte veranschaulicht werden, wie vielfältig und variabel der Gebrauch von Werbeslogans als verfestigten Sätzen in außerwerblichen Kontexten ist sowie wie und wozu Sprecher/-innen sie verwenden. Dadurch versuchten wir nicht nur zu verdeutlichen, dass Werbeslogans in der Gemeinsprache ähnlich wie Sprichwörter funktionieren und verwendet werden, sondern dass die Werbesprache eine moderne Quelle für die Entstehung neuer Sprichwörter darstellt. Es kann beobachtet werden, dass es in der Sprachgemeinschaft ein ausgeprägtes Bewusstsein von Sätzen als funktional vollständigen Einheiten gibt. Ob diese satzwertigen Gefüge als *Sprichwort*, *Slogan*, *Motto*, *Devise* u. a. bezeichnet werden, scheint für die Alltagssprecher/-innen unwichtig, was metakommunikative Elemente im Kotext von Slogans sowie Sprichwörtern veranschaulichen. Damit ließ sich auch die lexikografische Beschreibung von Werbeslogans nach demselben Modell wie dem für Sprichwörter sowie ihre Integration in das *OWID-Sprichwörterbuch* begründen. Stellenweise wurden die Slogan-Artikel jedoch im Hinblick auf die Spezifik der Werbeslogans erweitert („Kernform", „Äquivalente in anderen Sprachen", „Geschichte", „Vorkommen in Nachschlagewerken"), um auf die für Slogans wichtige Entstehungsgeschichte und ihre bisherige sporadische Dokumentation zu verweisen. Der künftige Schwerpunkt bei der lexikografischen Beschreibung des Korpusgebrauchs von Werbeslogans wird neben der Erarbeitung neuer Artikel auf der Weiterentwicklung der Methode der Erfassung lexikalischer Variationsmuster und von neuen Darstellungsformaten (z. B. in Form von Lückenfüllertabellen vgl. Steyer 2013) mit *lexpan* liegen.

Das *OWID*-Sprichwörterbuchmodul *Werbeslogans* richtet sich zum einen an In- und Auslandsgermanist/-innen und Deutschlehrende, die das Modul in ihre Lehrtätigkeit integrieren könnten, sowie an alle Deutschlernenden und Interessierten, die anhand zahlreicher Links das Thema Korpusgebrauch von Werbeslogans für sich entdecken wollen. Die korpusinformierte Untersuchung und Beschreibung von Werbeslogans soll aber auch für die Forschung anregend sein, insbesondere im Bereich der Sprichwortforschung und der Konstruktionsgrammatik, aber auch im Bereich der Werbesprachenforschung.

7.   Endnoten

1.   Es war die korpuslinguistische Wende, die das Erforschen und lexikografische Erfassen von sprachlichen Phänomenen im Allgemeinen und von Phraseologismen bzw. „usuellen Wortverbindungen" (Steyer 2013) im Speziellen tiefgreifend veränderte. Die korpuslinguistische

Wende führte im Bereich der Phraseologie zu zwei Paradigmenwechsen: Liberalisierung und Erweiterung der phraseologischen Einheiten in Richtung „funktionale Verfestigung" und musterbasierte Phraseologie (vgl. Steyer 2013).

2. Das *Wörterbuchportal OWID* enthält neben dem Sprichwörterbuch folgende lexikografische Ressourcen: *elexiko — Online-Wörterbuch zur deutschen Gegenwartssprache* (online seit 2013), *Paronymwörterbuch* (online seit 2018), *Kommunikationsverben* (online seit 2013), *Kleines Wörterbuch der Verlaufsformen im Deutschen* (online seit 2013), *Deutsches Fremdwörterbuch — Neubearbeitung* (Buchstaben A-H online seit 2016), *Neologismenwörterbuch* (online seit 2014), *Schulddiskurs 1945–55* (online seit 2008), *Protestdiskurs 1967/68* (online seit 2012), *Schlüsselwörter der Wendezeit 1989/90* (online seit 2015).

3. Die deutschen Wörterbuchartikel wurden im Rahmen des multilingualen EU-Projekts *Sprich-Wort. Eine Internetplatform für das Sprachenlernen* (2008–2010) erarbeitet und anschließend von Dr. Kathrin Steyer für die Online-Publikation zubereitet.

4. Bekannte Werbeslogans wurden beispielsweise im Slowenischen noch seltener lexikografisch erfasst: Einige Slogan-Artikel finden sich in einer kolaborativ geschriebenen Online-Ressource des gesprochenen Slowenisch „Razvezani jezik" (http://razvezanijezik.org/?page=Naslovnica), die punktuell die Entstehungsgeschichte und die Bedeutung skizzieren, ohne konkrete Sprachbelege anzuführen. Weitere lexikografische Ressourcen in anderen Sprachen sind mir nicht bekannt.

5. Wie in jüngsten korpusempirischen Arbeiten im Bereich der Phraseologie üblich, wird auch in der vorliegenden Untersuchung eine vordefinierte Liste von Slogans mit vermuteter Tendenz zur Usualisierung anhand eines virtuellen Korpus aus DeReKo validiert. Allerdings wird auch bei einem korpusbasierten („corpus-based") Vorgehen wie diesem immer wieder beobachtet, dass Erkenntnisse anhand Korpusevidenzen in den Forschungsprozess einfließen und ihn beeinflussen. Folglich wird die Festlegung auf ein Korpusparadigma in jüngster Zeit von Forschern relativiert: Der corpus-based- und corpus-driven-Ansatz schließen sich gegenseitig nicht aus und nur ihre Verknüpfung kann wirklich ertragreich sein (vgl. Steyer 2013: 71-72). Um dieser Relativierung terminologisch gerecht zu werden, wird in der vorliegenden Untersuchung in Anlehnung an Gredel (2014) von einem korpusinformierten Ansatz gesprochen.

6. Die Adaption der musterbasierten Korpusmethdologie für Slogans und ihre lexikografische Beschreibung im Modul *Werbeslogans* im *OWID-Sprichwörterbuch* wurden in Kooperation mit dem Projekt *Usuelle Wortverbindungen* (Leitung: Dr. Kathrin Steyer) während mehrerer Forschungsaufenthalte seit 2009 am Leibniz-Institut für Deutsche Sprache erarbeitet.

7. Die Suchprozedur basiert auf einer bestimmten Suchsyntax, dessen Komplexität von der Komplexität der Oberflächenform des Slogans abhängt. Diese wurde im Rahmen des EU-Projekts *SprichWort. Eine Internetplattform für das Sprachenlernen* sowie des UWV-Projekts entwickelt.

8. Ergebnisse der Suchanfragen werden zunächst in Form von KWIC-Ansicht (Kontextzeilen des Suchwortes/des Suchsyntagmas) analysiert. Die KWIC-Ansicht liefert einen ersten Überblick dazu, ob die Suchanfrage bereits eine hohe Anzahl an relevanten Treffern erbringt oder ob sie durch bestimmte Parameter anzupassen ist, d. h. ob die Suchanfrage zu eng oder zu weit ist. Zu jeder KWIC-Zeile kann man einen Volltext abrufen, der je nach Wahl einige (Ab-)Sätze vor und nach dem zentralen Suchsyntagma angibt.

9.   Das Analysewerkzeug wurde vom Projekt *Usuelle Wortverbindungen* (UWV) am Leibniz-Institut für Deutsche Sprache in Mannheim entwickelt.

10. Bevor mit dem Analyseprogramm *lexpan* eine KWIC-Systematisierung erfolgen kann, müssen zunächst anhand einer adäquaten Suchanfrage aus dem *DEREKO* entsprechende KWIC-Listen ermittelt werden. Diese werden in *lexpan* exportiert und darin abgespeichert sowie in einem nächsten Schritt mithilfe von einfachen Suchanfragen ohne komplexe Such-syntax systematisch und schnell untersucht.

8.   Literaturverzeichnis

Androutsopoulos, Jannis. 1997. Intertextualität in jugendkulturellen Textsorten. Klein, Josef und Ulla Fix (Eds.). 1997. *Textbeziehungen. Linguistische und literaturwissenschaftliche Beiträge zur Intertextualität*: 339-372. Tübingen: Stauffenburg.

Betz, Ruth. 2006. *Gesprochensprachliche Elemente in deutschen Zeitungen*. Radolfzell: Verlag für Gesprächsforschung.

Bubenhofer, Noah. 2009. *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kultur-analyse.* Sprache und Wissen 4. Berlin/New York: Walter de Gruyter.

Büchmann, Georg. 1864, ⁴²2001. *Geflügelte Worte: der klassische Zitatenschatz.* 42. neu bearb. u. aktual. Aufl. München: Ullstein Verlag.

Burger, Harald. 2015. *Phraseologie. Eine Einführung am Beispiel des Deutschen.* 5. Aufl. Berlin: Erich Schmidt Verlag.

Burger, Harald, Annelies Buhofer und Ambros Sialm. 1982. *Handbuch der Phraseologie.* Berlin/New York: Walter de Gruyter.

Cosmas-II (Projekt „Cosmas II" (2015). Korpusrecherche- und -analysesystem. Institut für Deutsche Sprache. Mannheim). http://www.ids-mannheim.de/cosmas2/ [12.1.2021].

DEREKO (Institut für Deutsche Sprache (2016): Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2016-I (Release vom 31.03.2016)). Mannheim: Leibniz-Institut für Deutsche Sprache. www.ids-mannheim.de/DeReKo [12.1.2021].

Duden 12: Zitate und Aussprüche. 2002, ³2008. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

Ďurčo, Peter, Kathrin Steyer und Katrin Hein. 2017. *Sprichwörter im Gebrauch.* Unveränderter Wiederabdruck der 2015 in Trnava erschienenen Erstausgabe. Mannheim: Institut für Deutsche Sprache.

elexiko (Online-Wörterbuch zur deutschen Gegenwartssprache). http://www.owid.de/wb/elexiko/start. html [12.1.2021].

Engelberg, Stefan, Annette Klosa-Kückelhaus und Carolin Müller-Spitzer. 2019. Lexikographie zwischen Grimm und Google? *Sprachreport* 35(2): 30-34.

Fix, Ulla. 1997. Kanon und Auflösung des Kanons. Typologische Intertextualität — ein ‚post-modernes' Stilmittel? Eine thesenhafte Darstellung. Antos, Gerd und Heike Tietz (Eds.). 1997. *Die Zukunft der Textlinguistik. Traditionen, Transformationen, Trends:* 97-108. Reihe Germanis-tische Linguistik 188. Tübingen: Niemeyer.

Fix, Ulla. 2007. Der Spruch — Slogans und andere Spruchtextsorten. Burger, Harald et al. (Eds.). 2007. *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung:* 1. Halbband: 459-468. Handbücher für Sprach- und Kommunikationswissenschaft 28(1). Berlin/New York: Walter de Gruyter.

Gredel, Eva. 2014. *Diskursdynamiken. Metaphorische Muster zum Diskursobjekt Virus.* Berlin: Mouton de Gruyter.

Hars, Wolfgang. 2002. *Das Lexikon der Werbesprüche. 500 bekannte deutsche Werbeslogans und ihre Geschichte.* Frankfurt a. Main: Piper Verlag.

Hemmi, Andrea. 1994. *„Es muß wirksam werben, wer will nicht verderben": Kontrastive Analyse von Phraseologismen in Anzeigen-, Radio- und Fernsehwerbung.* Zürcher Germanistische Studien. Frankfurt a. M.: Peter Lang.

Janich, Nina. 1997. Wenn Werbung mit Werbung Werbung macht … Ein Beitrag zur Intertextualität. *Muttersprache* 107: 297-309.

Janich, Nina. 2010. *Werbesprache. Ein Arbeitsbuch.* 5. erw. Aufl. Tübingen: Narr Studienbücher.

Janich, Nina. 2019. Intertextualität und Text(sorten)vernetzung. Janich, Nina (Ed.). 2019. *Textlinguistik. 15 Einführungen und eine Diskussion.* 2. Aufl. 177-198. Tübingen: Narr.

Jeromin, Rolf. 1969 (o. J.). *Zitatenschatz der Werbung. Slogans erobern Märkte.* Gütersloh: Präsentverlag Heinz Peter.

Jesenšek, Vida. 2011. Sprichwörter im Wörterbuch. *Linguistik online* 47(3): 67-78. http://www.linguistik-online.de/47_11/.

Jesenšek, Vida. 2013. Das lexikographische Beispiel in der Parömiographie. Formen und Funktionen. *Lexikos* 23: 150-171. http://lexikos.journals.ac.za/pub/article/view/1209/720.

Keibel, Holger. 2008. *Mathematische Häufigkeitsmaße in der Korpuslinguistik: Eigenschaften und Verwendung.* Mannheim: Institut für Deutsche Sprache. URL: http://www.ids-mannheim.de/kl/dokumente/freqMeasures.html [13.07.2020].

Klosa, Annette und Carolin Müller-Spitzer unter Mitarbeit von Martin Loder. 2016. *Internetlexikografie. Ein Kompendium.* Berlin/Boston: Walter de Gruyter.

Klosa-Kückelhaus, Annette und Carolin Müller-Spitzer. 2019. OWID und OWID[plus]: lexikographische und lexikalische Ressourcen am IDS Mannheim. *Zeitschrift für germanistische Linguistik* 47(2): 418-431. Berlin/Boston: Walter de Gruyter.

lexpan — Lexical Pattern Analyzer (Version 2019-06-21). Ein Analysewerkzeug zur Untersuchung syntagmatischer Strukturen auf der Basis von Korpusdaten. Entwickelt vom Projekt "Usuelle Wortverbindungen", Institut für Deutsche Sprache, Mannheim. http://uwv.ids-mannheim.de/lexpan/ [13.1.2020].

Liste geflügelter Worte. https://de.wikipedia.org/wiki/Liste_gefl%C3%BCgelter_Worte [12.1.2020].

Lüger, Heinz-Helmut. 1998. Vom Zitat zur Adaption. Zu einigen Verwendungsweisen satzwertiger Phraseologismen. *Beiträge zur Fremdsprachenvermittlung* 34: 118-135.

Lüger, Heinz-Helmut. 1999. *Satzwertige Phraseologismen. Eine pragmalinguistische Untersuchung.* Wien: Ed. Praesens.

OWID (Online-Wortschatz-Informationssystem Deutsch des IDS). https://www.owid.de/ [12.1.2021].

Polajnar, Janja. 2012. Textuelle Aspekte von rekontextualisierten Werbeslogans in deutschsprachigen Zeitungen. Eine korpusbasierte Untersuchung bekannter Werbeslogans im Zeitungskorpus des Deutschen Referenzkorpus (*DeReKo*). *Muttersprache* 122(1): 48-64.

Polajnar, Janja. 2016. Recontextualisation of International Advertising Slogans and their Equivalents in Different European Languages. *Poznań Studies in Contemporary Linguistics* [Online ed.] 52(1): 85-117.

Polajnar, Janja. 2019. *Werbeslogans im aktuellen Sprachgebrauch. Eine korpusinformierte, diachrone Untersuchung zur Dynamik des Slogan-Gebrauchs mit lexikografischen Fallstudien.* Amades 55. Mannheim: IdS Leibniz-Institut für Deutsche Sprache.

Razvezani jezik. Prosti slovar žive slovenščine. http://razvezanijezik.org/?page=Naslovnica [12.3.2021]

Redensarten-Index (Wörterbuch für Redensarten, Redewendungen, idiomatische Ausdrücke und feste Wortverbindungen). http://www.redensarten-index.de/suche.php [12.1.2021].

Schlobinski, Peter. 1989. Frau Meier hat Aids, Herr Tropfmann hat Herpes, was wollen Sie einsetzen? Exemplarische Analyse eines Sprechstils. Schlobinski, Peter et al. (Eds.). 1989. *OBST 16. Sprache und Erfahrung*: 1-34. Osnabrücker Beiträge zur Sprachtheorie 41. Osnabrück: Univ. Osnabrück.

Slogans.de. http://www.slogans.de/slogans.php?Op=SRanking1 [12.1.2021].

Steyer, Kathrin (Ed.). 2012. *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie.* Studien zur Deutschen Sprache 60. Tübingen: Gunter Narr Verlag.

Steyer, Kathrin. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht.* Studien zur Deutschen Sprache 65. Tübingen: Narr.

Steyer, Kathrin (Hg.). 2018. *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen.* Studien zur Deutschen Sprache 79. Tübingen: Narr.

Steyer, Kathrin und Peter Ďurčo. 2013. Ein korpusbasiertes Beschreibungsmodell für die elektronische Sprichwortlexikografie. Benayoun, Jean-Michel, Natalie Kübler und Jean-Philippe Zouogbo (Eds.). 2013. *Parémiologie. Proverbes et formes voisines.* Band 3: 219-250. Sainte Gemme: PUSG.

Steyer, Kathrin und Janja Polajnar. 2015. Werbeslogans. Sprichwörterbuch in OWID. http://www.owid.de/wb/sprw/start.html.

SWB: Sprichwörterbuch in OWID. http://www.owid.de/wb/sprw/start.html [12.1.2020].

SWB-SL: **Steyer, Kathrin und Janja Polajnar.** 2015. Modul *Werbeslogans* im OWID-Sprichwörterbuch. https://www.owid.de/service/stichwortlisten/slgn [12.1.2020].

SWP (EU-Sprichwortplattform). http://www.sprichwort-plattform.org/ [12.1.2020].

Towards a Legal Dictionary Dutch–Limburgish: Preferences and Opportunities

Conrad J.P. van Laer, *Limburgish Academy Foundation, Maastricht, The Netherlands (coenvanlaer9@gmail.com)*

Abstract: Limburgish is a regional language in the Limburg region consisting mostly of the Belgian and Dutch provinces of Limburg. The discussion will be on the Dutch province of Limburg where Limburgish has received some legal recognition as a language. It is desirable to have a Dutch–Limburgish legal dictionary to communicate Dutch law, although Limburgish legal terminology is not a working legal language in any legal system.

Special structures and data are needed to collect accessible contents for a Dutch–Limburgish legal dictionary. However, there is no blueprint to create a bilingual legal dictionary for a single jurisdiction. Therefore, this paper first considers some options for the design of the intended dictionary before going into possible sources for Dutch and Limburgish data.

The concluding remarks deal with the immediate challenges to be overcome before reaching the production stage of a useful and fully-fledged Dutch–Limburgish legal dictionary.

Keywords: SPECIALIZED DICTIONARIES, SUBJECT FIELD DICTIONARY, LEXICOGRAPHY, MULTILINGUAL DICTIONARIES, BILINGUAL DICTIONARIES, EQUIVALENCE, CULTURE-DEPENDENT DOMAINS, LEGAL LANGUAGE, DUTCH, REGIONAL LANGUAGE, LIMBURGISH

Opsomming: Op weg na 'n regswoordeboek Nederlands–Limburgies: Voorkeure en geleenthede. Limburgies is 'n streektaal in die Limburg-streek wat meestal uit die Belgiese en Nederlandse provinsies van Limburg bestaan. In die bespreking word gefokus op die Nederlandse provinsie van Limburg waar Limburgies reeds 'n bietjie wetlike erkenning as taal ontvang het. Alhoewel die Limburgiese regsterminologie nie 'n funksionerende regstaal in enige regstelsel is nie, is dit wenslik om oor 'n Nederlands–Limburgiese regswoordeboek te beskik om die Nederlandse reg weer te gee.

Spesiale strukture en data word benodig om toeganklike inhoud vir 'n Nederlands–Limburgiese regswoordeboek te versamel. Daar is egter geen konsep vir die skep van 'n tweetalige woordeboek vir 'n enkele jurisdiksie nie. Daarom word die opsies vir die ontwerp van die beplande woordeboek eers in hierdie artikel beskou voordat moontlike bronne vir Nederlandse en Limburgiese data bespreek word.

Ter afsluiting word daar gekyk na die onmiddellike uitdagings wat oorkom moet word voordat 'n bruikbare en volledige Nederlands–Limburgiese regswoordeboek saamgestel kan word.

Sleutelwoorde: GESPESIALISEERDE WOORDEBOEKE, VAKWOORDEBOEK, LEKSIKOGRAFIE, MEERTALIGE WOORDEBOEKE, TWEETALIGE WOORDEBOEKE, EKWIVALENSIE,

KULTUURAFHANKLIKE DOMEINE, REGSTAAL, NEDERLANDS, STREEKTAAL, LIMBURGIES

1.   Introduction

Limburgish is a regional or minority[1] language spoken in the Limburg region, which consists of the Belgian and Dutch provinces of Limburg. Linguistically, Limburgish differs from the neighbouring languages Dutch and German because of its own phonology, grammar, and vocabulary.[2] The recognition the language received in 1997 under Part II of the European Charter of Regional or Minority Languages (ECRML)[3] is confined to the Dutch province of Limburg. The focus in this paper will therefore be on Limburgish in the South-Eastern province of Limburg in the Netherlands unless otherwise specified.

Since lawyers often need to explain the meaning of legal terminology to clients during consultations, it may be assumed that Limburgish is used in conversations about legal matters. Moreover, as part of the recognition of Limburgish under Part II of the ECRML, in particular under Article 7(1)(d), the Dutch government is currently obliged to facilitate and/or encourage the use of Limburgish, in speech and writing, in public and private life. Public life covers the domains of government and the courts.[4] In the longer term, recognition of Limburgish under part III could take place. A part III recognition would include more robust measures for the language to be used in the administration of justice (De Groot 2019: 121-122).

Currently, Limburgish is not used in any jurisdiction. The Dutch province of Limburg is fully dependent on the Dutch jurisdiction which means that Dutch is the only official language of government for this province. Limburgish is not a legal language and strictly speaking legal sources in Limburgish do not exist. Nevertheless, useful translations of Dutch legal terms may be found in non-legal Limburgish sources. These Limburgish equivalents most likely do not cover all Dutch legal terms, but a Dutch–Limburgish legal dictionary could help to give Limburgish access to the legal domain. The intended Dutch–Limburgish legal dictionary will expand the usage of Limburgish in the legal sphere and will fulfil the prevailing obligation under part II ECRML.

However, to date there is no Dutch–Limburgish legal dictionary which could help legal professionals to inform their clients about Dutch law in Limburgish. A Dutch–Limburgish legal dictionary could help to make current Dutch law more comprehensible, thereby contributing to the propagation of Limburgish legal terms.

A bilingual dictionary is a tool pairing Dutch legal terminology with Limburgish legal terminology. Both parts of the dictionary must be presented in such a way that they are useful for the prospective users i.e., the target group of the dictionary. This paper takes account of that reality in its treatment of the special structure of the dictionary and the possible sources for the Dutch part and the Limburgish part of the dictionary. In addition to setting out preferences for the possible design and contents of a Dutch–Limburgish legal diction-

ary, this paper assesses the opportunities for completing such a labour-intensive process successfully.

2.   Design of the dictionary structure

A legal dictionary is not a simple word list in alphabetical order but primarily a linguistic tool consisting of many elements, for example Dutch entries or entry words, definitions, Limburgish equivalents, grammar, syntax, collocations, examples and explanatory sentences. Obviously, Dutch entries and their Limburgish equivalents are the compulsory elements of each dictionary article. Then, the compulsory and facultative elements will have to be structured to shape the whole bilingual dictionary article and to make the dictionary useful. However, there is no blueprint available to produce a Dutch–Limburgish legal dictionary. This means that the design of the structure of the intended dictionary is an important step. To this end, determining the target group of intended users of the dictionary is important as the design of the dictionary must correspond with the needs of those users.

First of all, the target users and their needs must be identified. Professional lawyers who want to discuss Dutch law and their Limburgish speaking clients will be the main target groups of the intended legal dictionary. The intended audience is bilingual since monolingual speakers are rare in modern Limburg. It is estimated that 70–75 percent of a population of about one million in Dutch Limburg speaks Limburgish.[5] Although most people in Dutch Limburg are able to speak both Dutch and Limburgish, when it comes to justifying themselves before a court of law, they may feel the need to express themselves in Limburgish, the language which is emotionally closest to them or in which they have greater fluency.

Today, documents in Limburgish do not have legal effect so lawyers may not look to the intended dictionary to help with the drafting in Limburgish, but rather to communicate with laypeople preferring Limburgish, particularly, to explain Dutch legal concepts to those clients. Consistently, the intended dictionary does not so much aim to produce texts in Limburgish but to improve the understanding of Dutch legal terminology used in conversations in Limburgish. Therefore, the intended dictionary does not have to include any Limburgish translations of Dutch legal definitions as these are too-complicated-to-explain Dutch terms. In addition, it is not necessary to prescribe any uniform terminology, for example by excluding some synonyms or by neglecting language or spelling variants. The main function of the dictionary is communicative as it aims to help laypeople to understand Dutch law in Limburgish without having to decipher legalese.

As there is no existing bilingual legal Dutch–Limburgish dictionary we have no models to guide us, but there is an interesting bilingual legal dictionary for another regional language in the Netherlands. The 'Juridysk Wurdboek' is a Dutch–Frisian legal dictionary, which was published two decades ago for an

audience of lawyers and civil servants using legalese professionally (Duijff 2000: 11). This reference work will provide inspiration for the design of the intended Dutch–Limburgish legal dictionary. It must be noted though that Frisian is an official language of government while Limburgish does not share that status yet.

The 'Juridysk Wurdboek' is a bilingual dictionary that is unidirectional, allowing translations from Dutch to Frisian. The index from Frisian to Dutch is a simple word list in alphabetical order that refers the users back to the dictionary entries to ensure that they can easily identify the source terms. This macrostructure could be a starting point for the arrangement of the entries or entry words in the intended Dutch–Limburgish legal dictionary.[6]

The microstructure of the 'Juridysk Wurdboek' could be a starting point as well. The microstructure of a dictionary describes the arrangement of the information provided by the dictionary articles. Obviously, the macrostructure must be aligned to the microstructure of the envisaged dictionary so the index of the 'Juridysk Wurdboek' reflects the dictionary entries. These entries consist of one word only. Compound terms, for example the Dutch term 'buitenvervolgingstelling', are single words and can be accepted as entries. It is noteworthy that to facilitate user access, multi-word expressions for legal concepts are not included as separate entries but may be added to the main entries. Multiword terms that are not featuring as separate entries have the result that dictionary articles frequently refer to separate entries that may include fixed expressions and example sentences as well (Duijff 2000: 13-15).

Given that Limburgish is not on an equal footing with the Frisian language, the dictionary must be compiled for oral communication with laypeople in the first place. Spoken legal language makes the insertion of pronunciation aids more important than it would be in dictionaries for authors of formal documents. The provision of pronunciation guidance will prevent mistakes made by users in oral communication, especially as far as lower frequency terms are concerned (Tihelková 2006: 117). Preferably, the intended dictionary including the pronunciation guidance will be digital.[7] Digital voice output will be more user-friendly for lawyers and laypeople than the International Phonetic Alphabet (IPA) in a printed dictionary. Ideally, bilingual conversations on legal matters will be translated almost real-time on a phone or tablet.[8]

It is difficult, sometimes, if not impossible to dislodge established usage. This is perhaps the main reason that most lexicographers acknowledge that dictionaries must register actual word usage. Accordingly, the Dutch–Frisian dictionary presents the contemporary Frisian legal language except where an exception had to be made and a neologism had to be derived from Old Frisian sources.[9] The second exception to the accepted principle of describing actual word usage is that the dictionary does not list synonyms for Frisian terminology. The reason for this choice is that the dictionary aims to be prescriptive to avoid any ambiguity of written terms and to achieve a uniform legal terminol-

ogy (Duijff 2000:19-20). However, this consideration does not apply to Limburgish legal terms which have no claim to any legal consequences.

3.   Sources for Dutch legal terms

The design of the Dutch–Limburgish legal dictionary must be tailored to the needs of the target group of professional lawyers, such as judges or civil servants, who want to explain Dutch law to laypeople who prefer Limburgish. Presumably, professional lawyers are aware of most if not all definitions of Dutch legal terms, so accurate and verifiable Dutch definitions are only needed to ascertain that these dictionary users have found the correct term. Definitions are not needed to assess degrees of equivalence between Dutch terms and Limburgish terms since that equivalence is always full: conceptual distinctions are absent as both languages relate to the same Dutch legal system, whose only binding legal terminology is in the Dutch language.

Arguably, most Limburgish translations of legal definitions would be useless in conversations as legal definitions are too complicated to help laypeople to understand Dutch law. The main purpose of the Dutch–Limburgish dictionary is not so much the transfer of legal concepts and their specific definitions but the explanation of the use of Dutch legal terms by means of Limburgish translations. Preferably, these translations will feature in illustrating sentences, like the following ones: 'D'n avvekaot vroog de rechter um straofvermindering' (The lawyer asked the judge to grant a mitigation of a penalty); 'Volges de euvereinkoms höb iech aonspraok op ein daarde vaan de wins' (Pursuant to the agreement I have a claim on one third of profit); 'Dit ligk boete zien competentie' (This is beyond his competence); 'Krach vaan wèt verkriege' (To gain force of law).

To compile a comprehensive legal dictionary Dutch–Limburgish, Dutch legal terminology should be the starting point. Obviously, terms which are exclusively legal must be included in the intended dictionary as they must be explained in everyday language. In addition, common words like 'divorce' or 'contract' must be included since they can acquire a specific meaning in a legal context. Although it may be problematic to answer the question how legal terms can be distinguished from non-legal terms,[10] it may be necessary that sources for Dutch terminology cover both terms and common words used in a legal context.

Both terms and common words used in a legal context are available in a small number of legal dictionaries which are relevant sources for Dutch legal terms. The most recent edition of the Dutch legal dictionary, the title of which still includes the name of its founder (Van Caspel and Damen 2016), is an extensive source of current Dutch legal terminology covering a substantial number of collocations[11] in many legal domains. To clarify legal terms, the dictionary provides many definitions or explanations and it frequently refers to relevant legal provisions. In addition to this work, another comprehensive

Dutch legal dictionary exists (Khan 2018). Both dictionaries have not been based on any accepted or representative standard corpus of Dutch legal documents. A comparison of both sources should take place to assess their exact scope and academic quality.[12] This means that the contents of the intended dictionary must be derived from one or two traditional legal dictionaries for Dutch law which may be supplementary to some extent.

4.   Sources for Limburgish legal terms

As stated earlier, Limburgish legal terms cannot be found in legal documents since Limburgish is not an official language used in the legal domain in any jurisdiction. Currently, there is no Limburgish legal language as a source for legal terminology. Therefore, sources must be traced that will provide Limburgish terms that can be easily recognized and accepted as legal terms corresponding to Dutch equivalents.

The first source to be considered is the growing corpus of Limburgish texts collected by the Limburgish Academy.[13] The Limburgish Corpus reflects the written usage from 1775 to the present.[14] However, the Limburgish Corpus is not aligned and bilingual as it does not include similar legislative or judicial documents in both Dutch and Limburgish. Therefore, Dutch legal dictionaries of sufficient quality are needed to assess possible translations of Dutch legal terms from the Limburgish Corpus. From 2021, the process of mining this expanding Limburgish Corpus will be facilitated as metadata will provide access to different types of texts that can be mined for Limburgish terms with some legal connotation. Firstly, the corpus will be tagged for text topics, including 'legal'. This label is used for everyday speech referring to legal topics. Secondly, the corpus is tagged for 'genre' (nonfictional prose) and subgenre (like minutes of meetings, policy statements, and regulations). The use of the language in these more formal and administrative domains could also help to trace Limburgish terminology that has a more legal connotation.

The size of the Limburgish Corpus will likely suffice to provide translations for high frequency everyday terminology and to extract example sentences to illustrate their usage in speech referring to legal contexts. It is expected that the Limburgish Corpus will cover terminology such as 'echsjeiing' (divorce), 'besjikking' (decision), 'verdachde' (suspect) or 'perces-verbal' (report). These examples show an etymological relationship between Limburgish and Dutch terms, but this relationship is not necessary. The following examples suggest different origins: 'aonpakke' (bring charges against), 'beveurmunde' (have custody over), 'proofrech' (right of judicial review), 'toch' (usufruct) and 'tous' (lease term). More importantly, Limburgish terms may be embedded in multi-word expressions that will include collocations, like 'Zie make 't gesjèl aonhengeg bij de rechbaank' (They submit the dispute before the court).

Novels and other secondary sources in the Limburgish Corpus will document colloquial speech, but it is less certain whether it will include enough low

frequency terms, which are exclusively legal, for example 'comparant' (person appearing before the court etc.) or 'obligatoir' (obligatory). In addition, compound terms like the Dutch 'buitenvervolgingstelling' or 'ingebrekestelling' are terms illustrating the gap between everyday language and legal language (Van den Bergh 1979: 54). Therefore, it will be assumed that the Limburgish Corpus will not provide enough terms to translate all Dutch terms that are present in the standard legal dictionaries.

Additional methods are needed to create Limburgish legal terms which are low frequency terms that do not feature in everyday language and that are not attested in the Limburgish Corpus. Other solutions found in previous research (De Groot 2012: 541-544)[15] will be discussed here: preserving the original term, paraphrasing, and creating neologisms.

Preserving the original Dutch term does not provide a Limburgish equivalent revealing anything about the meaning of that Dutch term. Preserving the Dutch term would not make it more familiar or easier for Limburgish speakers to understand it. This rejection of the first solution includes 'transliterations' assimilating the phonology of a Dutch legal term into the phonological system of Limburgish (Stephens and Boyce 2014: 305) to make borrowings more like the native words of this target language. For example, transcribing 'descente' as 'dessânte' is no translation and will not be very helpful to understand the legal concept involved.

Paraphrasing is the second solution and consists of describing Dutch terms using different words. Paraphrasing is driven by precision, and the descriptive method might help to explain the essential characteristics of the concept (Fuglinszky and Somssich 2020: 760). This approach may be effective, provided the intended meaning has been spelled out in neutral language and terms have been omitted (Šarčević 2000: 252). However, a paraphrase consisting of several common words could be too long or too complex to make it easy to understand the exact meaning of the original Dutch legal term. This means that paraphrasing should be limited.

The third subsidiary solution, the neologism,[16] looks more promising. Neologism means the artificial creation of a new term in the target language that would obviously not have a fixed meaning therein, and so it is perfectly suited to express the special nature of the concept of the source legal system, which lacks approximate equivalents in the target legal language (Fuglinszky and Somssich 2020: 762). Although neologisms are not in use in Limburgish yet, they could prove to provide useful new terms if they reveal the original Dutch term to some extent. By contrast, invented words for arbitrary sounds like 'blubs' have no legal connotation, are not transparent and fail to provide any information.

Finally, candidate equivalents might be found in Limburgish documents dating from the period before 1600 when Limburgish was used as a legal language of government.[17] However, accurately transcribed legal documents up to 1600 need to be evaluated by Middle Dutch and Middle Limburgish linguis-

tic experts to ascertain the text's linguistic provenance. This may be an expensive, time-consuming procedure. Another caveat for using possible equivalents from older texts is that most professional lawyers will not be familiar with Limburgish terms before 1600 even though to some extent the use of archaisms and obsolete words is typical of legal language. In this case most terms before 1600 will refer to outdated legal systems or local situations. Linguistically, these pre-1600 terms will likely not be recognizable for Limburgish-speaking laypeople.[18]

The intended dictionary will not be prescriptive and will therefore include synonyms. It is certainly possible that more Limburgish candidates could be deemed viable options to translate Dutch legal terms. However, the intended dictionary should provide Limburgish terms which can be easily recognized as legal terms translating Dutch equivalents. In those cases, the primary selection criterion will be that the Limburgish terms are memorable and self-explanatory to facilitate their spread and their usage. Suitability of Limburgish terms according to this criterion will be assessed based on two requirements. Firstly, the frequency of the usage of the term in the Limburgish Corpus. Frequent usage of a word will likely facilitate its recognition and its acceptance by a lay audience and possibly a professional audience. Secondly, if feasible, a user group of lawyers and paralegals will be consulted, via questionnaires, about the suitability and acceptability of proposed Limburgish legal translations for Dutch legal terminology.

Finally, attention should be paid to one possible complicating factor in creating a Limburgish legal terminology. Limburgish is not only used in the Dutch province of Limburg but also in the province of Limburg under Belgian sovereignty, where Dutch is also used as a language of government. It cannot be excluded that at some point another Dutch–Limburgish legal dictionary will be developed for the Dutch legal terms used by the Belgian jurisdiction. However, legal terms are system-bound because they are related to and embedded in legal systems. For this reason, the dictionary compiler translating Dutch legal terms from the Netherlands into Limburgish should try to avoid similarities with Belgian legal terms. This requires that the choice of Limburgish terms for the jurisdiction in the Netherlands must be based on a comparison of Dutch and Belgian law in order to express the differences between both legal systems. For example, the Limburgish translation of the Dutch term used in the Netherlands 'officier van justitie' (public prosecutor) should differ from the one for its Belgian counterpart 'procureur des konings' (Crown prosecutor) (Knap-Dlouhá and Škrlantová 2008: 86-87). These terms belong to two legal systems using Dutch as a legal language. Introducing only one Limburgish equivalent would cover up the differences involved.

5.   Feasibility: preferences and opportunities

It is unrealistic to plan a comprehensive and explanatory legal dictionary

Dutch–Limburgish to be finished in less than five, maybe ten years unless some financial resources are made available to a team of professionals. Therefore, one should consider a less ambitious pilot project to test a provisional work-flow,[19] to develop clear instructions for the dictionary compilers, to specify the data and the structures needed for the intended dictionary, and to get feedback from users in order to improve the initial data and structures. The pilot testing should be limited to one legal subdomain within larger areas, for example criminal law, allowing decisions to be made for the full-fledged production phase that is compliant with high quality standards.

The pilot project should focus on a limited number of primary Dutch sources, in particular dictionaries, to test whether they are compatible with the microstructure of the intended dictionary. This compatibility could be a prob-lem since many multi-word entries appear in Van Caspel and Damen 2016, the main source for Dutch legal terms, while multi-word terms are absent in the entries of the 'Juridysk Wurdboek', the dictionary which provides the inspira-tion for the design of the intended Dutch–Limburgish legal dictionary. In addi-tion, it should be assessed which Dutch terminology is relevant and sufficient not only to extract some Limburgish equivalents and illustrating sentences from the corpus of Limburgish texts, but also to show for which Dutch entries translations are missing.

For possible Limburgish terminology, the pilot project should explore to what extent possible candidates, multi-word expressions and collocations can be found in the Limburgish Corpus.[20] This exploration is important as it has been argued above that it is almost impossible to create Limburgish terms which are exclusively legal but not appearing in everyday language. If the Limburgish Corpus Dictionary provides only a very limited number of Lim-burgish candidate equivalents for legal terms, it could become difficult to maintain that the intended dictionary will help legal professionals to inform their clients about Dutch law in Limburgish. To fill this gap, time-consuming paraphrasing will be needed since preserving the original Dutch term is not an option while creating acceptable neologisms will be complicated.

Apart from the data needed for the intended dictionary, attention should be paid to its compilation and, after its completion, its publication and dis-semination. Publication in book form is an option, as is making it available online. For an online version, the software tools may provide pronunciation guidance as a facultative element of the dictionary. By contrast, an accessible search interface for the end users is a compulsory element. Ideally, the diction-ary must be comprehensible enough to answer the basic queries of a non-expert while the legal practitioners must feel confidence in using its transla-tions (Poon 2010: 90). Preferably, to reach both target groups and to get their feedback, the Limburgish Corpus Dictionary will be expanded to include the Limburgish Legal Dictionary.[21] Incorporating a full-fledged legal terminology in the Limburgish Legal Dictionary would be an important step towards law-

yers drafting documents in Limburgish thereby anticipating the enhanced protection of this regional language (De Groot 2019: 119-125).

Acknowledgements

Endnotes

1. The term "minoritized" language may be used in preference to "minority" because it draws attention to the unequal power relationships between languages. Limburgish does not refer to any regional variation of Dutch spoken in Dutch or Belgian Limburg.
2. Limburgish Language: https://limburgs.org/en/limburgish/
3. ECRML:
    https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680695175
4. Cf. Explanatory Report to the European Charter for Regional or Minority Languages Strasbourg, 5.XI.1992 https://rm.coe.int/16800cb5e5 p. 10: (62) "The charter does not lay down precise objectives in this respect but is content to call for an effort of promotion." However, the Dutch government has been negligent in its implementation of Part II obligations and no request for recognition under Part III seems to be forthcoming.
5. Contemporary presence: https://limburgs.org/en/limburgish/
6. An additional Dutch–English index could be derived from (Van den End 2016).
7. Dictionaries are made available in digital format for several reasons. Digital editions eliminate the cost of publishing print dictionaries and allow updates to the dictionaries' content. Thirdly, publishing online and in apps allows users instant access to dictionaries on devices they carry everywhere on their person, making them much more convenient and portable. See e.g., Andrews and Prys 2016: 10.
8. Cf. Google's Translate app which suggests this is technically possible, for many languages: https://support.google.com/translate/answer/6142474?hl=en&ref_topic=7011659 Cf. the interactive online tool produced by the Living Tongues Institute for Endangered Languages. This device digitally preserves words and phrases, and allows the user to hear high-quality audio recordings of their language, as well as record and upload new content and images: https://livingtongues.org/talking-dictionaries/
9. Old-Frisian legal terms refer to concepts from a legal system that is completely different from the modern Dutch system. In addition, Old-Frisian legal terms are outdated (Duijff 2000: 17).
10. There is no clear boundary. To a certain extent, it can be argued that any matter may be the subject of legal rules but obviously everyday events such as buying a loaf of bread are not essentially legal acts. To find out whether the context is legal, it is problematic to steer a mid-

dle course, but it may help that the context explicitly refers to situations involving disputes or lawyers.

11. An analysis of collocations in Dutch legal language has been carried out on the dictionary 'Fockema Andreae's juridisch woordenboek' (Tryczyńska 2014).

12. Accepted or representative standard lists of basic or most frequently used Dutch legal terms are not available. The Justice Thesaurus offers standard keywords that may be useful to identify Dutch legal terminology, but this Thesaurus is not complete: e.g., 'obligatoir' is missing. More Dutch legal terms could be found in selected legislative documents, as listed in the Introduction of the 'Juridysk Wurdboek' (Duijff 2000: 20-22).

13. https://limburgs.org/en/corpus/ Within the Limburgish Corpus the contemporary spelling of the Maastricht variety is linked to all other existing spelling variation in other Limburgish dialects.

14. See https://limburgs.org/en/corpus/ and Michielsen-Tallman et al. 2017.

15. It is possible to distinguish more techniques: Kozanecka et al. 2017: 88.

16. Seven criteria to assess neologisms have been listed: Mac Aodha 2018: 295.

17. Prof. Dr. Louis Berkvens stated that after 1600 more foreign elements were included in Limburgish legal sources.
    The Meuse-Rhineland region was already fragmented politically in the Middle Ages and was divided even further from 1600 due to raids from neighbouring countries. The new rulers introduced their languages for administration and justice in their Limburgish territories. Depending on the occupying power, French, Dutch, or High German were used as languages of government: https://limburgs.org/en/limburgish/

18. For example, 'bauduyn', a legal term before 1500 (for 'fine' in English), could be deemed an obscure term; cf. De Maasgouw 1881, p. 428 second column. Available:
    https://resolver.kb.nl/resolve?urn=dts:2394003:mpeg21:0004

19. Cf. Chiocchetti et al. 2013: 9-11: The terminology workflow.

20. Abbreviations could be excluded from translating into Limburgish. Cf. Duijff 2000: 15.

21. https://limburgs.org/en/dictionary/ Filters for legal domains could make the build-in legal dictionary accessible as a separate unit. These filters must be based upon a predefined classification using an established division of the legal domain into subdomains such as criminal law etc.

References

A.   Dictionaries

Duijff, P. 2000. *Juridisch woordenboek Nederlands–Fries met een index Fries–Nederlands.* Groningen/ Leeuwarden: Fryske Akademy/Martinus Nijhoff. Available:
https://taalweb.frl/wurdboekportaal

Khan, W. 2018. *Groot juridisch woordenboek: Beschrijvende definities van termen binnen de Nederlandse rechtspraktijk; inclusief een Engelse vertaling van de termen en verwijzingen naar relevante jurisprudentie.* Den Haag: Sdu Uitgevers.

Van Caspel, R.D.J. and M.P. Damen. 2016. *Fockema Andreae's juridisch woordenboek.* 16th revised edition. Groningen: Wolters-Noordhoff. Also published online:
https://www.vandale.nl/fockema-andreaes-juridisch-woordenboek-nu-online

Van den End, A. 2016. *Juridisch–Economisch lexicon NL–EN. The Legal and Economic Lexicon.* 6th edition. Deventer: Wolters-Kluwer. Available:
https://www.gatewaywoordenboeken.nl/Start-159.html

B.   Online databases

Justice Thesaurus. 2019. Research and Documentation Centre: Knowledge Centre in the Field of the Dutch Ministry of Justice and Security, version 2019. Available:
https://english.wodc.nl/documents/regulations/2020/11/08/justice-thesaurus-version-2019
or in Dutch: Justitiethesaurus, versie 2019. Available:
https://www.wodc.nl/documenten/regelingen/2020/11/18/justitiethesaurus-versie-2019

C.   Other literature

Andrews, T. and G. Prys. 2016. Terminology Standardization in Education and the Construction of Resources: The Welsh Experience. *Education Sciences* 6(2): 1-15. Available:
https://www.mdpi.com/2227-7102/6/1/2

Chiocchetti, E., N. Ralli, V. Lušicky and T. Wissik. 2013. Spanning Bridges between Theory and Practice: Terminology Workflow in the Legal and Administrative Domain. *Comparative Legilinguistics* 16: 7-22. Available:
https://pressto.amu.edu.pl/index.php/cl/article/view/6427/6449

De Groot, G-R. 2012. Legal Translation. Smits, J.M. (Ed.). 2012. *Elgar Encyclopedia of Comparative Law:* 538-549. 2nd edition. Cheltenham: Edward Elgar.

De Groot, G-R. 2019. European Charter for Regional and Minority Languages. Schoonheim, T. and J. van Hoorde (Eds.). 2019. *Language Variation: A Factor of Increasing Complexity and a Challenge for Language Policy within Europe. Contributions to the EFNIL Conference 2018 in Amsterdam: 115-128. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences.* Available:
http://efnil.org/documents/conference-publications/amsterdam-2018/EFNIL2018-18-DeGroot.pdf

Fuglinszky, Á. and R. Somssich. 2020. Language-bound Terms — Term-bound Languages: The Difficulties of Translating a National Civil Code into a Lingua Franca. *International Journal for the Semiotics of Law* 33: 749-770. Available:
https://link.springer.com/article/10.1007%2Fs11196-020-09704-x

Knap-Dlouhá, P. and M. Škrlantová. 2008. Rechtstaallandschap: Intralinguale vertaling van juridische terminologie bij de didactisering van juridische vertaling. Hrnčířová, Z., E. Krol, K. Mercks, J. Pekelder en J. Ultzen: *Praagse perspectieven 5: Handelingen van het Regionaal Colloquium Neerlandicum van Midden-Europa aan de Karelsuniversiteit te Praag op 20, 21 en 22 September 2007*: 79-90. Praag: Universitaire Pers. Available:
https://www.dbnl.org/tekst/_pra004200801_01/colofon.php

Kozanecka, P., A. Matulewska and P. Trzaskawka. 2017. *Methodology for Interlingual Comparison of Legal Terminology: Towards General Legilinguistic Translatology.* Poznań: Contact. Available:
https://repozytorium.amu.edu.pl/bitstream/10593/21240/1/Dissertationes%2011%202017%20Kozanecka%20Matulewska%20Trzaskawka.pdf

Mac Aodha, M. 2018. *Lexicographie, traduction et langues minoritaires: le cas de l'irlandais au sein de l'Union européenne.* Doctoral Dissertation. Strasbourg: University of Strasbourg. Available: https://tel.archives-ouvertes.fr/tel-02081060/document

Michielsen-Tallman, Y., L. Lugli and M. Schuler. 2017. A Limburgish Corpus Dictionary. Digital Solutions for the Lexicography of a Non-standardized Regional Language. Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2017 Conference, Leiden, the Netherlands, 19–21 September 2017:* 355-376. Brno: Lexical Computing. Available: https://elex.link/elex2017/wp-content/uploads/2017/09/paper22.pdf

Poon, W.Y.E. 2010. Strategies for Creating a Bilingual Legal Dictionary. *International Journal of Lexicography* 23(1): 83-103

Šarčević, S. 2000. *New Approach to Legal Translation.* Reprint. The Hague: Kluwer Law International.

Stephens, M. and M. Boyce. 2014. The Struggle for Civic Space Between a Minority Legal Language and a Dominant Legal Language: The Case of Māori and English. Mac Aodha, M. (Ed.). 2014. *Legal Lexicography: A Comparative Perspective:* 289-319. Farnham: Ashgate.

Tihelková, A. 2006. *Theory and Practice of Czech Bilingual Specialized Dictionaries.* Doctoral Dissertation. Prague: Charles University.

Tryczyńska, K. 2014. Collocaties in het juridisch woordenboek van Fockema Andreae: Een frequentieanalyse. *Neerlandica Wratislaniensia* 24: 93-106. Available: https://wuwr.pl/nwr/article/view/3037/2954

Van den Bergh, G.C.J.J. 1979. *Recht en taal, preadvies.* Deventer: Kluwer.

The Yin and Yang of Dictionary Making: Slovenian Lexicographers Reflect

Alenka Vrbinc, *School of Economics and Business, University of Ljubljana, Slovenia (alenka.vrbinc@ef.uni-lj.si),*

Donna M.T.Cr. Farina, *Department of Multicultural Education, New Jersey City University, New Jersey, USA (dfarina@njcu.edu)*
and
Marjeta Vrbinc, *Faculty of Arts, University of Ljubljana, Slovenia (marjeta.vrbinc@ff.uni-lj.si)*

Abstract: This contribution presents the second and final report on a study that set out to gain a greater understanding of what general modern Slovenian lexicography is. The full study focused on the philosophy, accomplishments, daily practice, and dictionary projects of seven prominent members of the Slovenian lexicographic community, all of whom were interviewed at length. An open-ended interview script allowed study participants to both reflect on their practice and portray their vision of what lexicography or terminography is. Part 1 (Vrbinc, Farina and Vrbinc 2018a; 2018b) reported on whether the lexicographers saw their work as drudgery and what they saw as the nature of their role in the society. Here, in Part 2, the emphasis is on the day-to-day concerns of the participating Slovenian lexicographers. They discuss both the technological and the human resources that drive their work. This second report also includes a brief survey of the historical development of Slovenian lexicography as the necessary backdrop for the interpretation of the modern situation.

Keywords: INTERVIEW, LEXICOGRAPHER, LEXICOGRAPHIC PHILOSOPHY, LEXICO-GRAPHIC PRINCIPLES, LEXICOGRAPHIC THEORY, LEXICOGRAPHIC PRACTICE, SLO-VENIAN LEXICOGRAPHY, HUMAN RESOURCES, TECHNOLOGICAL RESOURCES, CROWD-SOURCING

Opsomming: Die negatiewe en positiewe van woordeboekmaak: Beskou-ings van Sloweense leksikograwe. Hierdie bydrae stel die tweede en finale verslag van 'n studie wat daarop gemik is om 'n beter begrip van die algemene moderne Sloweense leksikografie te verkry, bekend. Die volledige studie het gefokus op die filosofie, prestasies, daaglikse praktyk, en woordeboekprojekte van sewe prominente lede van die Sloweense gemeenskap met wie almal indringende onderhoude gevoer is. 'n Oop onderhoud het deelnemers aan die studie toegelaat om na te dink oor wat hulle doen en om ook weer te gee wat hul visie van die leksikografie of terminografie is. Deel 1 (Vrbinc, Farina en Vrbinc 2018a; 2018b) het verslag gedoen oor of leksikograwe hul werk as sleurwerk ervaar en wat hulle as die aard van hul rol in die gemeenskap beskou. Hier, in deel 2, is die klem op die daaglikse probleme van die deelnemende Sloweense leksikograwe. Hulle

bespreek beide die tegnologiese en menslike hulpbronne wat as dryfkrag vir hul werk dien. As noodsaaklike agtergrond vir die interpretasie van die moderne situasie bevat die tweede verslag ook 'n kort oorsig van die historiese ontwikkeling van die Sloweense leksikografie.

Sleutelwoorde: ONDERHOUD, LEKSIKOGRAAF, LEKSIKOGRAFIESE FILOSOFIE, LEKSI-KOGRAFIESE BEGINSELS, LEKSIKOGRAFIESE TEORIE, LEKSIKOGRAFIESE PRAKTYK, SLO-WEENSE LEKSIKOGRAFIE, MENSLIKE HULPBRONNE, TEGNOLOGIESE HULPBRONNE, SKAREBENUTTING

1.   Introduction

To a certain degree, dictionaries are created and delivered in similar ways worldwide. Bilateral and multilateral lexicographic work takes place between organizations (such as AFRILEX, ASIALEX, DSNA, and EURALEX) and between academies of science (such as the Austrian or Slovenian academies). While some lexicographers network through conference attendance and keep up with the scholarly literature, others are not able to follow every new development in the field. The present discussion is the second report based on an interview study of lexicographers. Its intended audience is any lexicographer, hard-pressed for time, who might wish to know more about what other lexicographers do and who might benefit from such information in daily practice. Before the full interview study was conducted, the authors were (and are still) not aware of the existence of any studies of practicing lexicographers. In the previous (Vrbinc, Farina and Vrbinc 2018a; 2018b) and in the present report, Slovenian lexicography is accessed via the perspectives of seven prominent lexicographers. Other researchers are encouraged to replicate this study in their own countries, to advance lexicographic theory as well as practice globally.

During February–March 2017, seven Slovenian lexicographers were interviewed over sixteen hours. Four overarching questions drove the full research study and the creation of an open-ended interview script:

A.   What is the philosophical and intellectual framework governing the work of Slovenian lexicographers? What ideas do they all share — across different institutions and projects — as they engage in making dictionaries?

B.   What are the main areas of concern and common significant problems that inform the work of Slovenian lexicographers?

C.   What do the lexicographers consider both the main strengths and the weaknesses of their current efforts in dictionary creation? What would they most like to change about their practice?

D.   What are the differences among our interviewees in their conception of what lexicography is all about?

The first report on the study's findings investigated the philosophy of the interviewees and focused on Research Question A above, with some elements of Research Question D. It explored views on prescription versus description, drudgery, and the role of the lexicographer as "harmless" or not (Vrbinc, Farina and Vrbinc 2018a; 2018b). This second and final report has an entirely different focus. It seeks answers to Research Questions B and C by investigating the seven lexicographers' work challenges and constraints, the major strengths of their current work situations, and the changes that they would like to see in future dictionary work. Note that only seven subjects were interviewed. From the outset, the full study was never intended to present an exhaustive picture of Slovenian lexicography; the study's findings might be different if more interviews were held or if different people were interviewed. Nevertheless, since the lexicographers interviewed all work on different projects and in different institutional settings (or independently), the present report as well as Report 1 does claim to provide important information on some of the pressing issues in lexicography today.

The present report is relevant to dictionary makers around the globe interested in improving their local lexicographic practice. By listening to the Slovenian lexicographers, we might all gain insight into what future practices could be most beneficial for our field. This report is also relevant to Slovenian theoreticians and practitioners of lexicography who wish to understand better the place of their work in the social context prevailing today — and the place of their work within the full history of Slovenian lexicography.

2.   A short history of Slovenian lexicography

As indicated above, the main objective of this report is to examine the work situation of Slovenian lexicographers today (Research Questions B and C above). In the case of Slovenia, to interpret today it is necessary to inspect yesterday, at least briefly. Speakers of the Slovenian language have known rule under the Habsburg Monarchy (1282/1335–1918), under the Austro-Hungarian Empire (1867–1918), and as part of the former Yugoslavia (1918–1991); Slovenia has been an independent country only since 1991. While by 1775 all children were required to attend school on the territory of modern Slovenia, the full use of the Slovenian language as the main medium of instruction at all educational levels is a relatively new development. In the second half of the nineteenth century, German was still the language of instruction for high school and university-level education. This state of affairs impacted the development of the Slovenian language as well as Slovenian lexicography.

Early literacy in Slovenia was connected with religion and dates to the late sixteenth century. In 1584, the first Slovenian word list was included in Dalmatin's Protestant Bible (*Biblia, tu ie, vse Svetu pismu ...*), the first translation of the Bible into Slovenian. The first print dictionary to include Slovenian words is Megiser's *Dictionarium quatuor linguarum ...* (1592); it has headwords in Ger-

man, followed by Latin, Slovenian, and Italian equivalents. A later dictionary by Megiser, the *Thesaurus polyglottus* (1603), included many more languages, among these: French, German, Greek, Hebrew, Hungarian, Latin, Polish, Slovenian, and Spanish.

These early Slovenian word lists and dictionaries with very simple equivalents gradually transitioned into a full-fledged bilingual lexicography, where more than simple equivalence was provided between a single pair of languages rather than a series of them. One hundred years later, by the end of the seventeenth century, high-quality bilingual dictionaries with Slovenian appeared, such as Ožbalt Gutsman's *Deutsch–windisches Wörterbuch* (in German, *windisch* is an old word for Slovenian) of 1789. Two well-respected dictionaries from the late nineteenth century with German and Slovenian are Maks Pleteršnik's *Slovenian–German Dictionary* [*Slovensko–nemški slovar*] (two vols., 1894–1895) and Matej Cigale's *German–Slovenian Dictionary* [*Deutsch–slowenisches Wörterbuch*] (1860). The first bilingual Slovenian dictionaries with English were intended for Slovenian immigrants to the United States: Košutnik's *A Pocket Slovenian–English and English–Slovenian Dictionary: Intended for Immigrants to America* [*Ročni slovensko–angleški in angleško–slovenski slovar: zlasti namenjen za izseljence v Ameriko*] (Ljubljana) of 1904 and Kubelka's *Slovenian–English Pocket Dictionary,* [*Slovensko–angleški žepni rečnik*], published in New York the same year.

No monolingual Slovenian dictionary existed until late in the twentieth century. Merše (2009) explains that before World War II, there was no special institute for the study of Slovenian. The Slovenian Academy of Sciences and Arts itself was established only in 1938, just before World War II. Within the Academy, the Institute of the Slovenian Language was established in 1945; today this is called the Fran Ramovš Institute after its first director.[1] Only in the 1950s did work begin on what would become the first monolingual Slovenian dictionary. In 1970, the first volume of *The Dictionary of Standard Slovenian* [*Slovar slovenskega knjižnega jezika*] (letters A to H) was published; all five volumes had appeared by 1991. In 1997, the full dictionary was reprinted in a single volume. The two-volume second edition appeared in print in 2014; it is now available online. Following an intense period of discussion and planning, work began in 2016 on the *e-Dictionary of Standard Slovenian* (eDSS) [eSSKJ]. While this new dictionary is informally called the "third" edition, it is conceived as an entirely new work rather than a revision of the earlier editions. It is being delivered to the public in once-per-year increments on FRAN,[2] the online dictionary portal of the Academy of Sciences (https://fran.si/). Approximately 500 entries are posted each year, with today's total at 1641. Current plans envisage that a print edition will be made available when the dictionary is complete.

Apart from monolingual lexicographic work taking place within the Academy of Sciences, from the mid to the late twentieth century there was an active independent bilingual lexicography in Slovenia; many dictionaries appeared from a variety of publishing houses. Apart from the existence of several small English and Slovenian dictionaries for immigrants to the US, starting in

the late nineteenth century and continuing up to the end of World War II (for example, Košutnik 1904 and Kubelka 1904), the pairing of Slovenian with English is a relatively late phenomenon. Škerlj's *English–Slovenian Dictionary* [*Angleško–slovenski slovar*], did not appear until 1944; it arose out of communication needs anticipated by the intellectual community during World War II (Rener 2007). The first monodirectional Slovenian–English dictionary, Kotnik's *Slovensko–angleški slovar,* appeared in 1945. More comprehensive coverage came much later, with the *English–Slovenian dictionary* [*Angleško–slovenski slovar*] of Grad et al. (1973). A larger work was published in 1978 (Grad et al., *Comprehensive English–Slovenian Dictionary* [*Veliki angleško–slovenski slovar*]). Krek's *Oxford–DZS Comprehensive English–Slovenian Dictionary* [*Veliki angleško–slovenski slovar Oxford*] appeared in the new century, 2005–2006, as did the *Angleško–slovenski slovar* of Vrbinc and Vrbinc (2009).

Due to close cultural ties, a pairing with the German language has been consistent throughout the Slovenian lexicographic tradition. In modern times, the Debenjak family's (Doris, Primož and Božidar Debenjak) *Comprehensive German–Slovenian Dictionary* [*Veliki nemško–slovenski slovar*] first appeared in 1992 and has remained in print, with the last publication in 1999. Their *Comprehensive Slovenian–German Dictionary* [*Veliki slovensko–nemški slovar*] of 1995 is likewise used to this day. After Slovenia gained independence from the former Yugoslavia in 1991, bilingual lexicography became more varied and comprehensive. There now exist recent dictionaries of Slovenian with Slavic languages (Czech, Polish, Russian, and Serbo-Croatian) as well as with Dutch, French, Italian, and Spanish. Unfortunately, just as the public's need for bilingual lexicographic tools has increased, Slovenian publishing houses have ceased to publish such print dictionaries.

Slovenian bilingual lexicographic work takes place outside of the Slovenian Academy of Sciences, which focuses on monolingual lexicography. Today in the Academy there are dictionaries or ongoing projects on orthography, synonymy, phraseology, and terminology (to name a few). However, monolingual work is not limited to Academy projects. In 2004, Trojina, Institute for Applied Slovene Studies, was founded in Ljubljana to promote linguistic research and the linguistic competence of Slovenian speakers. Trojina collaborates on projects with other institutions or academic units engaging in lexicographic work, at the University of Ljubljana and beyond.[3]

3.   The Slovenian study participants

Within the short time frame allotted for this study, we sought to represent as many locations as possible where dictionaries are being made in Slovenia, and to cover a variety of types of modern dictionary work. Only specialists who work on synchronic topics and concentrate on the standard language (as well as terminology) were interviewed. The Fran Ramovš Institute of the Slovenian Language in the Academy of Sciences and Arts is a center of dictionary work in

Slovenia. In the Lexicological Section, three of our interviewees concentrate on different aspects of lexicography: Nataša Jakop, Nina Ledinek, and Jerica Snoj. In another area of the Ramovš Institute, Mojca Žagar Karer has worked on numerous terminological dictionaries and directs the Terminological Section.

Apart from the Fran Ramovš Institute, there are ongoing lexicographic projects in a variety of units at the University of Ljubljana. There are also projects led by Trojina, Institute for Applied Slovene Studies, usually in cooperation with other units. We interviewed Apolonija Gantar and Iztok Kosem, researchers in the Faculty of Arts at the University of Ljubljana.

Lexicography in Slovenia does not only take place via institutionally-supported projects. Anita Srebnik is an independent lexicographer who authored the *Slovenian–Dutch European Dictionary* [*Slovensko–nizozemski evropski slovar*] (2006) and the *Dutch–Slovenian Dictionary* [*Nizozemsko slovenski slovar = Nederlands Sloveens woordenboek*] (2007), intended for Slovenian learners of Dutch. She is also an instructor of Dutch in the Faculty of Arts at the University of Ljubljana.

Due to their positions and influence, the reflections of these seven study participants are not anonymous; here, they are cited by name in order to advance lexicographic practice. The seven interviewees were given the option to speak "off the record," to make comments that would not be directly attributed to them. We received very few such comments; the participants were frank and forthcoming. For more detailed information on our subjects and how they were selected, see Vrbinc, Farina and Vrbinc (2018a), sections 7 and 8.

4.   The prevailing themes

Given that our full study sought to encourage future interview-based research on lexicographers worldwide, the interview script was published in its entirety (see Report 1, Vrbinc, Farina and Vrbinc 2018a; 2018b). While all of the interview questions informed Report 1 directly or indirectly, two of them, Script Questions 3a and 3b, were the main focus there. Three of the thirteen questions from the scripted interviews with the lexicographers serve as the main source of information here in Report 2; the answers to other questions were also consulted. Script Questions 10, 11, and 12 dealt with the challenges and obstacles the Slovenian lexicographers are confronted with as they try to deliver their products to users:

10. It goes without saying that lexicographic work takes place in the real world and is subject to the usual constraints and challenges of any practical work. In particular, there are always budgetary constraints, but not only budgetary. We would like to know: How is your work challenged by a variety of circumstances; what are the challenges and constraints?

11. Can you name the major strengths of your work situation? What is a best practice for you and your colleagues (e.g., access to different information/ sources, user-friendly dictionary-making software, cooperation with IT

specialists and/or corpus linguists and/or experts from other fields, etc.)? What affects most positively the compilation of your dictionaries?

12. If you could change one thing about the circumstances of your lexico-graphic work, what would it be? If you could change one feature of the lexicographic philosophy/theory that underpins your work, what would it be?

We identified holistically some themes that came up from the responses to these three questions; they are addressed below. Most often, the same or simi-lar themes arose in interviews with more than one person. We also included some ideas that can be considered relevant to our field even if they may not have come up repeatedly. Gouws (2012) provided a helpful framing for our discussion below. He asked: Who is a lexicographer? What does such a person do and what skill set should they have? Gouws identified four groups of peo-ple: (1) lexicographic practitioners without theoretical experience; (2) theoreti-cians without practical experience; (3) those who are primarily practitioners but have some theoretical background; and finally (4) theoreticians who have some practical experience (2012). Many aspects of Gouws' framing had parallels in the remarks of our study participants as they considered both the human and technological resources that contribute to successful dictionary projects.

4.1 Budget considerations

Our Question 10 above, while it mentions budget, was intended to steer the subjects toward addressing non-budgetary, non-financial constraints. Never-theless, every informant felt compelled to speak about the role that financial constraints play in lexicographic work. It was impossible for them not to speak of this, when budget affects almost every aspect of lexicography. Because this is already a familiar problem, their comments on budget are interwoven within the topics raised below.

4.2 Human resources

Given the technological nature of much modern dictionary work, it is common for the question of lexicographic resources to be considered as primarily one of hardware and software. However, our participants had many comments about the people in lexicography, the human resource.

4.2.1 Experts and expertise

The theme of experts and expertise arose often. Nataša Jakop of the Ramovš Institute brought up what she saw as a lack of the necessary expertise for lexi-

cography in Slovenia, a very small country (with a population of just above two million persons). She considered that a lexicographer needs to bring to the table an area of expertise, but not necessarily lexicography itself; this would correspond partially with Gouws' (2012) categories (3) or (4), practitioners with some theoretical background or theoreticians with some practical experience. According to Dr. Jakop, in order to be part of a dictionary-making team, a person needs training in phonology as well as other aspects of linguistics, corpus linguistics, information technology, and other areas. The ideal lexicographic team would have many different types of specialists on staff. Dr. Jakop noted that in a small country it is difficult to have all the necessary specialists on board for a single project; what is more, the budget for a large team is rarely available. This means that while technologically it would be possible to make a dictionary in five to ten years, without the budget for a sufficient number of experts, this is not likely to happen.

The terminographer among our study participants, Mojca Žagar Karer of the Ramovš Institute is satisfied with her experiences with specialists: She considers that working with the excellent and dedicated experts at the Ramovš Institute is a major advantage of her work situation. On the other hand, she encounters lack of motivation among the terminological field experts outside of the Ramovš Institute, who must constantly be encouraged to contribute. We might add: It is quite understandable that persons (such as the experts from other fields) who are not trained in lexicography would find the writing of terminological dictionary entries to be an onerous process.

Iztok Kosem, affiliated with Trojina, the Institute for Applied Slovene Studies, and the Faculty of Arts at the University of Ljubljana, alluded to the expertise of Slovenian lexicographers, who as a group are internationally well-connected and established. People from other countries want information from them, which can be considered a compliment. Recently, people from Portugal and Estonia traveled to Slovenia to see the methodology being used. They ended up improving upon the Slovenian technology for their target languages as well as upon the whole procedure.

As a form of expertise enhancement for themselves, Dr. Žagar Karer and Dr. Jakop would like more opportunities to attend conferences and workshops. Dr. Kosem pointed out that twenty years ago, lexicographers had to travel to get information, to talk to people. Now, more information can be obtained more easily for collaboration.

4.2.2   Cooperation and staffing

Most of the interviewees identified cooperation as a major strength in their work situations. Nataša Jakop noted that lexicographers in the Ramovš Institute can communicate with etymologists, orthographists, dialectologists, terminologists, and a large variety of other specialists. They can consult specialists in other fields (e.g., geography, anthropology), since the Ramovš Institute is

only one of eighteen institutes within the Academy of Sciences. Nina Ledinek in the Ramovš Institute noted how helpful it is to have so many specialists in different fields of linguistics, as well as experts in other areas of science employed at the Academy — there is always someone who can help with a problem. These experts have different backgrounds and specializations; they are young, enthusiastic, prepared, and eager to collaborate on problem-solving. Similarly, Mojca Žagar Karer emphasized her cooperative working environment, stating, "I am lucky with my colleagues."

Apolonija Gantar of the University of Ljubljana mentioned the cooperation between linguists and IT specialists (students and professors at the Faculty of Computer and Information Science who are interested in linguistic work). Jerica Snoj of the Ramovš Institute referred to a linguist/IT collaboration as the best experience of her lexicographic career. When her team began work on the *Dictionary of Slovenian Synonyms* in 2001, there was no appropriate program for writing and editing entries. The younger colleagues designed a program for this project, in line with the plan of the dictionary, that was completed in 2013. Dr. Snoj emphasized that it was the lexicographers/linguists who provided the programmers with very detailed instructions. She further noted that it is a very positive thing when lexicographers are supported by the appropriate technology. Iztok Kosem cited the value of international cooperation. There are numerous mutual projects and sometimes Slovenian lexicographers collaborate with foreign partners even when they do not receive funding. Dr. Kosem mentioned that he and his co-workers always strive to include people with knowledge of how to make information user-friendly and to have ongoing interactions with them.

Given the positive collaborations experienced by the study participants, it is not surprising that they expressed a desire for more collaboration — in the form of larger lexicographic staffs or teams. Nataša Jakop would like to have a bigger team working on phraseology (at the time of this study, three colleagues were working in this area); she considered that more specialists in lexicography should be working on every project. Dr. Jakop would also like to have more support staff (e.g., students) for some of the preliminary monotonous work, which is time-consuming and expensive. Similarly, Apolonija Gantar noted that there are levels of dictionary work that can be done by people with less expertise (fewer years of experience, trainees, students, crowd-sourcing). Nina Ledinek considered that hiring more people would allow for faster production as well as more accurate dictionary products. Dr. Ledinek maintained that while disagreements among specialists will never go away, they might be diminished if an improved financial situation reduced competition; competition for projects can be counterproductive. Mojca Žagar Karer, the only terminographer among our interviewees, agreed with Drs. Jakop and Ledinek that more people should be working on every project. In the case of terminological dictionaries, this means more (non-lexicographic) experts in each subject/field are needed, as terminographers are generally overburdened.

4.2.3 Crowd-sourcing

Crowd-sourcing has been seen by some in our field, including some of our study participants, as a possible solution to the problem of understaffing in lexicography. In their examination of crowd-sourcing in German and English versions of *Wiktionary*, Wolfer and Müller-Spitzer (2016) express skepticism that any solely crowd-sourced dictionary could reach the levels of quality seen in dictionaries created by professionals. Likewise, Sajous, Josselin-Leray, and Hathout (2018) see crowd-sourcing of neologisms (as examined in four dictionaries) as something complementary to but not replacing professional lexicographic work. Among our study participants, Apolonija Gantar mentioned crowd-sourcing in conjunction with discussion of budgetary problems. She considered that the distribution of public monies in lexicography is problematic and that projects could be carried out more cheaply if crowd-sourcing were used. Her remarks advocate for crowd-sourcing to both supplement and accelerate lexicographic work. Wolfer and Müller-Spitzer (2016) mention another positive of crowd-sourcing: "... it is a pleasure to see that there is a language-interested community that works on dictionaries voluntarily. Is this not also a sign for the relevance of dictionaries?" (368). While not mentioning crowd-sourcing directly, Mojca Žagar Karer noted that she gets good, satisfying feedback from the users of the terminological dictionaries who use the Ramovš Institute's consulting service.

Iztok Kosem mentioned crowd-sourcing as a convenient supplement to the work of lexicographic specialists. If users participate in contributing to a dictionary, it allows the team of experts to focus on more complex tasks such as defining or addressing phraseology and compounds. Dr. Kosem and his colleagues are trying to devise crowd-sourcing "tests" for people. For example, if volunteers are asked to attribute an illustrative example to a particular sense of a word, this gives the lexicographic team feedback as to whether their own division of senses is appropriate or whether an example should be linked to a given sense. Dr. Kosem noted that the dictionary user of today is more demanding: "Users are used to getting information instantly, participating, and having an opinion," and crowd-sourcing can favorably affect this dynamic by engaging users more. Dr. Kosem welcomes the new challenges caused by this type of user, "You cannot be proud of your dictionary if the user does not use it." He envisages each student coming to the dictionary with their own profile that the dictionary would adapt to. For example, a student of biology would see more illustrative examples in that field.

4.3   Technological resources

Our study participants expressed many opinions on the technological resources available to them for dictionary work. While they are glad to have modern tools available, they are not always fully satisfied with them and perceive that

their current tools are often not adequate to the tasks at hand. Anita Srebnik, an independent lexicographer, emphasized that software, different corpora and online dictionaries are more accessible now than they were in 2003, when she started her work on bilingual Slovenian–Dutch dictionaries. She considers support from language technology specialists to be absolutely essential in any lexicographic project.

Nina Ledinek of the Ramovš Institute would like better IT support. In addition, having access to different corpora would help her with her work: The corpus materials she is using now are ten or fifteen years old and do not contain enough Slovenian language material from the 1970s, 80s and 90s. Likewise, Iztok Kosem of Trojina and of the University of Ljubljana noted that the current Slovenian reference corpus (Gigafida) is 6 years out of date. Dr. Ledinek also mentioned that better dictionary software is needed; Mojca Žagar Karer made similar remarks. Jerica Snoj of the Ramovš Institute pointed out that not every corpus is useful for work on a dictionary; most of them were not initially designed for lexicography but for general linguistic investigations. For example, she would like to be able to look up full collocations or grammar points. Nataša Jakop of the Ramovš Institute would like to have a more automated way of selecting data and illustrative material. Finally, Dr. Snoj would like to change the belief that language technology alone is enough to make a modern dictionary without other linguistic knowledge; in her opinion, this is a dangerous idea that is growing in popularity.

The participants' discussion of technological resources overlapped with their discussion of human cooperation. Apolonija Gantar, researcher at the University of Ljubljana, believes that access to different information sources, user-friendly dictionary-making software, and cooperation with IT specialists and/or corpus linguists and/or experts from other fields all affect dictionary work positively. According to Dr. Gantar, people working at the Center for Language Resources and Technologies have all of this. Since every dictionary has a different underlying lexicographic concept, it goes without saying that the lexicographers must adjust to different projects; they have to get familiar with corpora and the information they can obtain in corpora.

Most interviewees noted how essential cooperation is between linguists and IT specialists, and some perceive a need for better cooperation. Iztok Kosem, a researcher at Trojina and the University of Ljubljana considers that the best lexicographic team has some people who know technology better and some "doubters" who challenge your ideas — whom you have to convince by providing explanations. In general, our participants' ideas are in line with Tarp (2012):

> The final conclusion is that lexicography will not cease to be an independent discipline with its own specific subject field as well as its own theory and practice but that it will tend to relate more and more to and interact with similar disciplines within the broad area of information science. ... [L]exicographical theory will place itself in an even better position to assist and guide the present transition from printed to electronic dictionaries. (329)

4.4   Theory and practice

Reflections on lexicographic theory found in the current literature have commonalities with the views of our distinguished interviewees. Anita Srebnik, an independent bilingual lexicographer, believes that current practices are ahead of theory; this is similar to Piotrowski (2013), who is skeptical of the validity of past theoretical models for our time, when rapid changes are taking place in both cultural and technological environments. Dr. Srebnik mentioned that new tools require the articulation of new theory as to how to use them; this recalls Gouws' (2018) exploration of guidelines for the adaptation of "different types of data distribution structures in online dictionaries" (178) as part of the transition from print to online. In line with Wiegand's famous 1984 pronouncement that lexicography is not a science and will not become one,[4] Srebnik considers theory and practice as two completely separate endeavors.

Apolonija Gantar, researcher at the University of Ljubljana, did not mention directly theory versus practice, but her remarks seem to speak to the separate "theoretical" statements that are imbued in each dictionary implicitly. Dr. Gantar believes that having more than one general monolingual dictionary of standard Slovenian would be good for Slovenia, because that would help people arrive at their own opinions about language; such multiple dictionaries as cultural artifacts would reflect a range of different user needs. This echoes Fuertes-Olivera and Tarp (2014), who note that dictionaries have covered "a wide range of different needs detected in society and … almost all spheres of human activity and knowledge" (39). Dr. Gantar maintained that dictionaries, as reference works intended for general users (who want to solve different linguistic problems) rather than for linguists, should allow room for users to doubt the very information they provide. Dictionary users should be able to rely on themselves and their own judgements. Users, in Dr. Gantar's opinion, should be more self-confident in their own linguistic knowledge. Another of our interviewees agreed with Dr. Gantar's desire for more than one general Slovenian dictionary, and emphasized that clashes in theory, rather than just competition for project money, drive the current Slovenian paradigm of a single large monolingual dictionary. This remark fits with the notion that a dictionary is itself a kind of theory; it possibly fits as well with Piotrowski's (2013) idea that there are multiple lexicographic theories.

Iztok Kosem, a researcher at several institutions including the University of Ljubljana, mentioned the necessity of engaging in multiple dictionary projects simultaneously, since a single project will only provide partial employment financially. Dr. Kosem noted that when a researcher writes a grant proposal, s/he must wait for the results of funding; some worthy projects are not funded at all (e.g., the comparative European project on dictionary use). It is not possible to focus, Dr. Kosem said, on a single project over a sequence of months or years, and when a researcher has too many projects it is difficult to master any one. Dr. Kosem would like to have not too many but more than a

single lexicographic project active at once. Luckily, he was able to connect work on a Slovenian–Hungarian dictionary and on a collocations dictionary in terms of reusing some skills and methods. While Dr. Kosem's remarks were intended to address the financial realities as well as the uncertainties of lexicographic work in the real world, they also have theoretical implications. A project that begins through a well-articulated grant proposal risks losing its theoretical focus as work is dragged out due to the vagaries of funding; moreover, very good projects that would advance lexicographic theory risk not coming to fruition.

Anita Srebnik, the only independent lexicographer in our group, to some extent echoed the concerns of Dr. Kosem, but with a different emphasis. In her case, she has non-lexicographic full-time employment (as an instructor of Dutch at the University of Ljubljana) that prevents her from being fully dedicated to lexicographic work. For this reason, it took her seven years to compile her bilingual Dutch–Slovenian dictionary. She found the organization of dictionary work to be very time-consuming and would have appreciated being part of a team that could have provided more support for her efforts. She came into lexicography without prior formal training and this also presented a challenge; for Gouws (2012), she would be considered in category (1), a practitioner who lacked theoretical experience. While such practitioners can succeed in creating outstanding products, Dr. Srebnik underlined that such conditions of lexicographic practice are far from ideal for lexicographic theory.

Nina Ledinek of the Fran Ramovš Institute enumerated several problems connected with everyday lexicographic work. Among other things, she considered that more material relevant to lexicography is needed on the standard language and language stratification. Lexicographers need to determine what is the standard language, what is dialectal, colloquial, etc. for dictionary work. While theory was not mentioned directly by this participant, it is clear that an important, if not *the* most important, element of lexicographic theory is the point of view of the lexicographer on the standard language (cf. Farina 2020).

Another problem Dr. Ledinek raised is that *Slovenian Orthography* [*Slovenski pravopis*], an important reference tool for the country,[5] is out of date since it was published in 2001 and users' habits have changed since then. This brings to mind Dr. Jakop's comment (above) about the desirability of completing dictionaries in five to ten years. When longer periods of time pass, a research tool may no longer be as effective: Either a work's underlying theory may cease to be suitable to modern reality or the theory underlying the work may lose focus due to the passage of time.

While some of the remarks above appear to indicate a recognition that there is no one lexicographic theory, nevertheless some interviewees do perceive that there is a prevailing general theory that guides their work. Mojca Žagar Karer of the Ramovš Institute maintains that there is no need to change lexicographic theory, but she would like to have more time to reflect on the theory and philosophy of her field, terminography. Dr. Gantar agrees with Dr. Žagar

Karer that the current state of lexicographic theory is acceptable, but adds that she is free to develop theory or change it if she wishes, through the publication of articles and the presentation of papers at conferences.

5.   Conclusions and recommendations

This second and final report on an interview study of seven prominent Slovenian lexicographers brings home the potential of the interview process to contribute to our understanding of modern lexicography. It would be impossible for our field to advance without the theoretical and practical contributions of the leading academic journals. In addition to these, the perspectives of the leading, working lexicographers "on the ground" provide an entirely different and valuable knowledge source. Just as dictionaries are cultural artifacts, embedded in the time and place of their creation, so lexicographers are cultural emissaries who both represent the users of the language and — as some study participants point out — collaborate or engage with users in the creation of or understanding of dictionary information. While these emissaries may be seen as "invisible" or "harmless" (see Vrbinc, Farina and Vrbinc 2018a) and may heretofore have been mostly ignored as subjects for scientific study, this is an oversight when much may be gained from listening to them. It is hoped that similar interviews will be conducted in other countries, to build a deeper comparative understanding of lexicographic ideas and work practices. It is worth considering whether the lexicographic product of a single nation can be evaluated fully without this type of grounded perspective from real lexicographers.

As a country, Slovenia is blessed with a powerful cadre of lexicographic experts, due in no small part to its successful programs for training future specialists (see Vrbinc, Farina and Vrbinc 2018a). For this reason (as one study participant points out), Slovenian lexicography has attracted international attention and has benefitted from international collaborations. This is a significant asset that could position Slovenia as a world leader in all things lexicographic. It is no small achievement for a country of only about two million people.

The Slovenian lexicographers interviewed expressed satisfaction with their work as well as with most aspects of their work situations. The study authors are left nevertheless with the impression of a lexicographic infrastructure that is inequitable. The interviews brought out that some Slovenian lexicographic work is independent and entirely uncompensated; some work is dependent on soft monies (i.e. grants) that might not be renewed, thus rendering some projects precarious; finally, some work is stable and supported institutionally on a permanent, ongoing basis. It is obvious that this situation may adversely affect the finances of the lexicographers themselves. Without the interviews, it would not be as clear how this state of affairs impacts the quality or the quantity of dictionary products, or how it affects theoretical innovation and long-term lexicographic development. Our participants are aware of the impact — though they are too busy being lexicographers to dwell on it. They

are aware of missed opportunities for dictionaries based on alternate theoretical premises, something that could potentially provide users with more choices and provide lexicography with more possibilities to develop as a discipline. Participants mentioned the need for fewer disagreements as well as less forced competition among specialists, the need to allow similar projects with different theoretical goals to coexist, and the need for better use of existing resources. Most likely, similar situations prevail in other countries. This is all the more reason to have interviews with specialists elsewhere, to provide us with that comparative picture.

Acknowledgements

Endnotes

1. Fran Ramovš (1890–1952) was a Slovenian linguist and a co-founder and member of the Slovenian Academy of Sciences and Arts. He chaired the Academy from 1950 to 1952. He published the *Historical Grammar of Slovenian* (Volume 2: Consonantism, 1924; Volume 7: Dialects, 1935), *Dialect Map of Slovenian* (1931), *A Short History of Slovenian* (1936), and *Slovenian Morphology* (1952).
2. The online portal FRAN is named after Fran Ramovš.
3. For more information on the setting of Slovenian lexicography and on the development and influences of Slovenian lexicographic theory, see Vrbinc, Farina and Vrbinc (2018a).

4.   "Lexicography was never a science, it is not a science, and it will probably not become a science. Scientific activities as a whole are aimed at producing theories, and precisely this is not true of lexicographical activities. We must bear in mind that writing on lexicography is part of meta-lexicography and that the theory of lexicography is not part of lexicography" (Wiegand 1984, 13).

5.   This two-part reference work consists of a section on the rules of Slovenian orthography and a dictionary section; it might more accurately be called a manual of Slovenian orthography.

References

Dictionaries

Bajec, Anton et al. 1970–1991. *Slovar slovenskega knjižnega jezika*. Ljubljana: Državna založba Slovenije.

Cigale, Matej. 1860. *Deutsch–slowenisches Wörterbuch*. 2 vols. Laibach: Herausgegeben auf Kosten des Hochwürdigen Herrn Fürstbischofs von Laibach, Anton Alois Wolf.

Debenjak, Doris, Božidar Debenjak and Primož Debenjak. 1992. *Veliki nemško–slovenski slovar*. Ljubljana: Državna založba Slovenije.

Debenjak, Doris, Božidar Debenjak and Primož Debenjak. 1995. *Veliki slovensko–nemški slovar*. Ljubljana: Državna založba Slovenije.

Grad, Anton, Ružena Škerlj and Nada Vitorovič. 1973. *Angleško–slovenski slovar*. Ljubljana: Državna založba Slovenije.

Grad, Anton, Ružena Škerlj and Nada Vitorovič. 1978. *Veliki angleško–slovenski slovar*. Ljubljana: Državna založba Slovenije.

Gutsman, Ožbalt. 1789. *Deutsch–windisches Wörterbuch: mit einer Sammlung der verdeutschten windischen Stammwörter, und einiger vorzüglichern abstammenden Wörter*. Klagenfurt: gedruckt und im Verlage bey Ignaz Aloys edlen von Kleinmayer …

Košutnik, Silvester. 1904. *Ročni slovensko–angleški in angleško–slovenski slovar: zlasti namenjen za izseljence v Ameriko*. Ljubljana: A. Turk.

Kotnik, Janko. 1945. *Slovensko–angleški slovar*. Ljubljana: Grad.

Krek, Simon et al. 2005–2006. *Veliki angleško–slovenski slovar Oxford*. Ljubljana: DZS.

Kubelka, Viktor J. 1904. *Slovensko–angleški žepni rečnik*. New York.

Megiser, Hieronymus. 1592. *Dictionarium quatuor linguarum*. A lohanne Fabro.

Megiser, Hieronymus. 1603. *Thesaurus polyglottus*. Frankfurt.

Pleteršnik, Maks. 1894–1895. *Slovensko–nemški slovar*. Ljubljana: Katoliška tiskarna.

Škerlj, Ružena. 1944. *Angleško–slovenski slovar*. Ljubljana: Grad.

Snoj, Jerica, Martin Ahlin, Branka Lazar and Zvonka Praznik. 2016. *Sinonimni slovar slovenskega jezika*. First edition. Ljubljana: Založba ZRC.

Srebnik, Anita. 2006. *Slovensko–nizozemski evropski slovar*. (Zbirka Evropski slovarji). Ljubljana: Cankarjeva založba.

Srebnik, Anita. 2007. *Nizozemsko slovenski slovar = Nederlands Sloveens woordenboek*. (Slovarji DZS). First edition. Ljubljana: DZS.

Toporišič, Jože, Franc Jakopin, Janko Moder, Janez Dular, Stane Suhadolnik, Janez Menart, Breda Pogorelec, Kajetan Gantar, Martin Ahlin and Milena Hajnšek-Holz (Eds.). 2001. *Slovenski pravopis*. Ljubljana: Založba ZRC.

Vrbinc, Alenka and Marjeta Vrbinc. 2009. *Angleško–slovenski slovar*. Ljubljana: Cankarjeva založba.

Other references

Čibej, Jaka, Vojko Gorjanc, Iztok Kosem and Simon Krek (Eds.). 2018. *Proceedings of the XVIII Euralex International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana*. Ljubljana: Ljubljana University Press, Faculty of Arts.

Farina, Donna M.T.Cr. 2020. 20th-century Soviet–Russian Dictionary Front Matter: Focus on the Linguistic Norm. *Lexicography* 7(1–2): 115-122.
Available: https://doi.org/10.1007/s40607-020-00064-z

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

Gouws, Rufus H. 2012. Who can Really be Called a Lexicographer? *Lexikos* 22: 217-225.
Available: http://lexikos.journals.ac.za/pub/article/view/1004 and https://doi.org/10.5788/ 22-1-1004

Gouws, Rufus H. 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke [A Lexicographic Data Pulling Structure for Online Dictionaries]. *Lexikos* 28: 177-195.
Available: https://doi.org/10.5788/28-1-1461

Hartmann, Reinhard Rudolf Karl (Ed.). 1984. *LEXeter' 83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*. Tübingen: Max Niemeyer.

Jackson, Howard (Ed.). 2013. *The Bloomsbury Companion to Lexicography*. London: Bloomsbury.

Merše, Majda. 2009. Slovensko zgodovinsko slovaropisje s konceptualno-razvojnega vidika. Stabej, Marko (Ed) 2009: 251-255.
Available: https://centerslo.si/wp-content/uploads/2015/10/28-Merse.pdf

Piotrowski, Tadeusz. 2013. A Theory of Lexicography: Is There One? Jackson, Howard (Ed.). 2013: 303-320.

Ramovš, Fran. 1924. *Historična gramatika slovenskega jezika: Konzonantizem II.* Ljubljana: Učiteljska tiskarna.

Ramovš, Fran. 1931. *Dialektološka karta slovenskega jezika*. Založila rektorat univerze kralja Aleksandra I. in J. Blasnika nasl., univerzitetna tiskarna v Ljubljani.

Ramovš, Fran. 1935. *Historična gramatika slovenskega jezika: Dialekti VII*. Ljubljana: Učiteljska tiskarna.

Ramovš, Fran. 1936. *Kratka zgodovina slovenskega jezika*. Ljubljana: Akademska založba.

Ramovš, Fran. 1952. *Morfologija slovenskega jezika*. Ljubljana: Državna založba Slovenije za Univerzitetno študijsko komisijo.

Rener, Tanja. 2007. Ružena Škerlj (1904–1992): svetovljanka, razumnica, (so)avtorica številnih izdaj angleško–slovenskega slovarja ter slovensko–češkega in češko–slovenskega slovarja. Šelih, Alenka et al. (Eds.). 2007: 404-406.

Sajous, Franck, Amélie Josselin-Leray and Nabil Hathout. 2018. Lexical and Semantic Neology in English: The Complementarity of Crowdsourced Dictionaries and Professional Dictionaries Viewed through the Filter of Neology. *Lexis. Journal in English Lexicology* 12: 1-34.

Šelih, Alenka, Milica Antić Gaber, Alenka Puhar, Tanja Rener, Rapa Šuklje and Marta Verginella (Eds.). 2007. *Pozabljena polovica: Portreti žensk 19. in 20. stoletja na Slovenskem.* Ljubljana: Založba Tuma and SAZU.

Stabej, Marko (Ed). 2009. *Infrastruktura slovenščine in slovenistike.* Obdobja, Symposium 28. Ljubljana: Znanstvena založba Filozofske fakultete.

Tarp, Sven. 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22: 321-332.
Available: https://doi.org/10.5788/22-1-1010

Vrbinc, Alenka, Donna M.T.Cr. Farina and Marjeta Vrbinc. 2018a. Objectivity, Prescription, Harmlessness, and Drudgery: Reflections of Lexicographers in Slovenia. *Lexikos* 28: 361-383.
Available: https://doi.org/10.5788/28-1-1469

Vrbinc, Alenka, Donna M.T.Cr. Farina and Marjeta Vrbinc. 2018b. Slovenian Lexicographers at Work. Čibej, Jaka, Vojko Gorjanc, Iztok Kosem and Simon Krek (Eds.). 2018: 199-208.

Wiegand, Herbert Ernst. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, Reinhard Rudolf Karl (Ed.). 1984: 13-30.

Wolfer, Sascha and Carolin Müller-Spitzer. 2016. How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions. *Lexikos* 26: 347-371.
Available: http://dx.doi.org/10.5788/26-1-1346

Problems in Defining Ethnicity Terms in Dictionaries

Dragica Žugić, *Faculty of Philology, University of Donja Gorica, Podgorica, Montenegro (dragica.zugic@udg.edu.me)*
and
Milica Vuković-Stamatović, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro (vmilica@ucg.ac.me)*

Abstract: Despite the fact that lexicographers have increasingly been taking more care when it comes to defining socially sensitive terms, we argue that ethnicity terms still remain rather poorly defined. In a number of online monolingual dictionaries we surveyed in this study, we find that ethnicity terms are generally simplistically defined, mostly in terms of geography and citizenship, and argue that such definitions are too reductionist and sometimes even erroneous. We also find that some disparaging ethnicity terms are not labelled as such in some of the dictionaries surveyed. We also present a case study from Montenegro, in which a dictionary of the national academy of sciences was immediately revoked over a few ethnicity and ethnicity-related terms, after a violent outcry from two of Montenegro's ethnic minorities, dissatisfied with how their ethnicities were defined and treated in the dictionary. Based on our survey and the earlier findings from the literature, we recommend that international dictionaries follow a standardised model of defining ethnicities, which would additionally refer to an ethnicity's culture and potentially language, and be as inclusive as possible. We also recommend that editors and lexicographers of national dictionaries pay special attention to how they define the ethnic terms relating to the minorities living in their country or region, following a combination of a standardised and a partly customised approach, which would take into account the specific features of the minorities.

Keywords: ETHNICITY TERMS, ETHNICITY-RELATED TERMS, DICTIONARIES, LABELS

Opsomming: Probleme met die definiëring van etniese terme in woordeboeke. Ondanks die feit dat leksikograwe in die definiëring van sosiaal-sensitiewe terme toenemend versigtigheid aan die dag lê, word hier aangevoer dat etniese terme steeds redelik swak gedefinieer is. In 'n aantal aanlyn eentalige woordeboeke wat ons in hierdie studie ondersoek het, vind ons dat etniese terme oor die algemeen simplisties gedefinieer word, meestal in terme van geografie en burgerskap, en ons redeneer dat hierdie definisies té reduksionisties en soms selfs foutief is. Dit het ook geblyk dat sommige neerhalende etniese terme in sommige van die woordeboeke wat ondersoek is nie as sodanig geëttiketteer is nie. Ons lê ook 'n gevallestudie uit Montenegro voor, waarin 'n woordeboek van die nasionale akademie van wetenskappe weens 'n paar etniese en etniesverwante terme onmiddellik onttrek is ná die geweldadige protes van twee etniese minderhede in Montenegro wat ontevrede was met die manier waarop hul etnisiteite in die woordeboek gedefinieer en hanteer is. Gegrond op ons ondersoek en die vroeëre bevindings uit die literatuur, stel ons voor dat internasionale woordeboeke 'n gestandaardiseerde model vir die definiëring van etniese terme volg wat ook sal verwys na die kultuur en moontlik taal van 'n etniese groep, en wat

so inklusief moontlik sal wees. Ons beveel ook aan dat redakteurs en leksikograwe van nasionale woordeboeke besondere aandag skenk aan die manier waarop hulle die etniese terme rakende minderhede wat in hul land of streek woon, definieer deur 'n kombinasie van 'n gestandaardi-seerde en gedeeltelik pasgemaakte benadering te volg waarin die spesifieke eienskappe van die minderhede in ag geneem sal word.

Sleutelwoorde: ETNIESE TERME, ETNIESVERWANTE TERME, WOORDEBOEKE, ETI-KETTE

1.   Introduction

A typical user expects a general dictionary to contain ethnicity terms and to define them (Rader 1989). However, even though lexicographers have been making great efforts to improve their treatment of various politically sensitive and socially charged terms over the last decades, ethnic terms in dictionaries are still described rather poorly, usually just in geographical terms and even in those cases, sometimes too restrictively. To some ethnic groups such definitions may be offensive and a dictionary can face a strong public backlash on account of this. In this paper we will describe one such recent case from Montenegro, as well as inspect the literature on the issue and comparatively analyse the defini-tions of ethnic terms in various online monolingual dictionaries (English, Ger-man, Italian, Croatian, Serbian, and Albanian).

We start the paper by surveying the relevant literature on how dictionar-ies treat politically sensitive terms, with a special focus on ethnic terms.

2.   Treatment of politically sensitive terms in dictionaries

The present paper follows the tradition of studying how dictionaries treat politically sensitive terms. Some of the studies exploring these issues include the following: the treatment of ethnic names (Rader 1989), racial terms (Murphy 1991, 1998) and political terms (Dieckmann 1975; Veisbergs 2002); ideological aspects of dictionaries in general (Moon 1989; Ezquerra 1995; Wierzbicka 1995); use of offensive language (Schutz 2002); the issue of political correctness, with special emphasis on the treatment of gender (Barnickel 1999), etc. As can be seen, these studies overlap somewhat in their topics of interest and how they approach them; for our present purposes, we will say that the present paper deals with how ethnic names or ethnonyms, as politically and socially charged terms, are dealt with in dictionaries.

As Murphy (1998) argues, this type of research is conducted with two purposes in mind. The first refers to highlighting the inaccuracies and preju-dice in dictionaries so that they can be corrected in later editions and prevented from occurring in new dictionaries. Insensitive treatment of some terms (par-ticularly racial and ethnic, Murphy notes) may provoke a public outcry, as well

as organised boycotts, protests and even the banning of a dictionary (see Sub-section 2.2). One such recent case will be described later in the paper (Section 4). The second purpose of this type of research is to make a contribution to how we understand the relations between language, on the one hand, and social attitudes and categorisations, on the other.

All the studies mentioned reveal that dictionary definitions are indeed some-times insensitive or ideologically charged. Reflecting the nature of human beings, the vocabulary of every language contains "unpleasant" language. Therefore, naturally, the lexicon itself will reflect unpleasant stereotypes (Schutz 2002: 640). All authors agree that ideologies, political and social aspects will always be present in the definitions and that ideologically neutral entries as a whole can-not exist (cf. Moon 1989; Ezquerra 1995; Schutz 2002). Veisbergs (2002) notes that even some seemingly innocuous choices, such as the choice of a spelling variant, for instance, may reflect ideology — e.g. by our subscribing to British spelling in this paper we take a position and reject the American one. Some of the other choices we make may be seen as offensive by some groups and the same definitions may be seen quite differently by different ideological groups.

Items in dictionaries may be offensive in two ways, Schutz (2002) finds — either directly, i.e. those used offensively with a deliberate intention, typically name-calling (e.g. *nigger* for a black person) or indirectly, which is far more often the case. For instance, in the examples accompanying the entry for the noun *research* in the online Cambridge English Dictionary[1], we found an over-use of the pronoun *he* vs. *she*. Namely, in the 7 examples accompanying the definition of the term, three contained a third person singular personal pro-noun and all with a male referent (e.g. *his researches …; he dedicated his life to science; he emphasised …*). While modern dictionaries are making great efforts to avoid such infelicities, they unavoidably still do happen. This is due to the fact that many modern dictionaries are based on authentic corpora and authentic cor-pora reflect reality, which, in this case, includes the reality that some occupa-tions are stereotypically seen as predominantly male jobs (typically those requiring physical strength, but also some highly intellectual ones, such as being a researcher). In addition, corpora themselves are composed of texts which are censored (every text, prior to its publication, undergoes some sort of censor-ship, at least self-censorship), which means that the word list based on such a corpus may not contain some words which are politically undesirable (Wierz-bicka 1995: 194).

Another issue frequently cited as leading to bias in dictionaries is the fact that lexicographers are just humans and thus have human weaknesses, such as their debts to other people and their attachment to certain ideas (Ezquerra 1995: 151). They are also very likely to be imbued with their own culture, which shapes their understanding and perceptions (Veisbergs 2002). This all, of course, applies to the editor(s) of a dictionary as well. Landau (1984: 303)[2] finds that dictionaries reflect prejudice and views of the upper classes, those established and well-educated ones, and so present what is valued by such groups. Dic-

tionaries thus often reflect the leading social ideologies (Kalogjera 2001: 263).

Context is also very important in the study of potentially insulting words. Murphy (1991: 21) notes that "much of the usage labelling of racial, sexual and other epithets is based on the assumption that a member of the outgroup is using the term to describe the given ingroup". And while *nigger* may be used by the ingroup members to refer to themselves, it is certainly derogatory when used by the outgroup.

These issues especially become visible in the dictionaries produced under totalitarian regimes. Veisbergs (2002) observes that, *inter alia*, in such dictionaries some words are typically banned, while others, usually the politically charged ones, are purposefully misrepresented; in addition, some political terms are given plenty of dictionary space whereas some easily undergo a U-turn revision after certain ideas are rehabilitated in the society concerned.

These are, of course, extreme cases of visible bias in a dictionary; however, the dictionaries produced in modern democratic societies also feature some ideological distortions, in a much milder form, such as the gender bias discussed above. In the modern world, there is a growing need to correct such issues in language in general and, consequently, in dictionaries. Much of this has been driven by the movement advocating *political correctness*, which started in the '70s (Barnickel 1999). Some of the corrections made in the English dictionaries on account of greater sensitivities include a different treatment of the compound occupational names containing a "man"-element, e.g. the dictionaries now tend to add an admonitory note to terms for occupations ending in *-man* (e.g. *salesman, policeman* …). The examples accompanying definitions in the dictionaries are now carefully chosen to avoid ones reflecting stereotypes; and, increasingly, care is taken to use neutral pronouns such as *I* or *they* or *everyone,* etc. instead of *he* and *she*; etc. (Barnickel 1999). The requests made in the name of political correctness have also been criticised as unduly exaggerated in some cases (which is why the term has itself deteriorated and now it might even have a negative ring to many), whereas the very concept is somewhat controversial and contended by those opposing any type of censorship (Busse 2000).

2.1   Treatment of ethnic terms in dictionaries

Before we delve into the research done on ethnic terms or ethnonyms, we will define what mean under the term "ethnicity". This complex term has received many different definitions in various social sciences, but most of them present it very broadly in scope and do not clearly distinguish between "ethnicity" and other close terms, such as "race" and "nationality". For Horowitz (1985) and many authors who follow his classification, ethnicity is, in fact, an umbrella term for these concepts (Chandra 2006). Similarly, some authors, including Francis (1947), Rothschild (1981), Connor (1984) and Brass (1991), do not systematically distinguish between an ethnic group and a nation, and find the two largely synonymous (Gabbert 2006). For them, an ethnicity may refer to a

minority in a state, e.g. French Canadians in Canada, as well as to the French in France (Gabbert 2006). We will adhere to such an understanding of the term ethnicity, covering both these situations. In this paper, the term ethnicity was chosen rather than the term nationality for two reasons, cited in Xu (2002): first, while *ethnicity* is more of an academic concept, *nationality* is rather a legal and/or political one; second, *ethnicity* as a term can be more widely used than the term *nationality*, i.e. *ethnicity* can be a synonym for *nationality*, while the converse is not always the case. Thus, under such a broad definition, all the examples cited in this paper are considered as ethnicity labels.

The research on ethnicity terms can be divided into two strands — on the one hand, many studies have dealt with how the insulting nature of some ethnicity words is labelled in dictionaries and, on the other, little research has been conducted into the definitions of the ethnic terms which are generally not offensive.

Busse (2000), for instance, studies the insulting abbreviations for ethnicities (*Frog, Jap, Kraut …*) in some English learners' dictionaries and argues that, even though these terms are racist, students need to know their connotations when encountering them and advises on including them in the dictionaries. The dictionaries Busse studied varied on whether and how many of these terms they included, as well as what labels were used for indicating the genre range in which the terms are used — some were marked as *taboo* and some as *informal* (which suggests less insulting connotations than *taboo*). Busse commends the fact that the *Oxford Advanced Learner's Dictionary* puts a ! sign next such terms, as an indication to a foreign learner that these words should be avoided in use.

In the same vein of research, Norri (2000) and Nissinen (2015) find that learners' dictionaries typically have more warnings for potentially insulting words, whereas slang dictionaries (Nissinen 2015) tend to use them the least. On the other hand, some dictionaries simply decide to omit such entries (Norri 2000). In this strand of research we also find a study conducted by Ştefănescu (2015) on labelling disparaging ethnicity words in Romanian dictionaries and Wachal's (2000) study of labelling of taboo words in British and American dictionaries.

When it comes to defining ethnicity terms that themselves have no such obvious insulting connotations, fewer studies are available. Rader (1989) argues that ethnic terms should be defined in dictionaries. Despite the typical derivational and etymological connection which exists between an ethnonym and a certain place-name, geographical criteria should not be the sole criteria in such definitions, he argues. He specifically mentions the problem of an "over- or under-lap between ethnic group and citizenship", typically ignored in the dictionaries. As an example, he gives the term *Hungarian*, usually defined as "a native/inhabitant of Hungary", even though there are other ethnicities living in Hungary (Serbs and Slovaks, among others) as well as significant communities of ethnic Hungarians living in other countries (Romania, Serbia, The Czech Repub-

lic, Slovakia, etc.). Rader (1989: 133) states that "in some instances ethnicity should be separated from citizenship, and each accorded a separate definition, though I hesitate as to where the line should be drawn".

In one part of her paper, Murphy (1998) deals with the ethnicity terms in South African English dictionaries. She notes that complex issues were sometimes oversimplified in the dictionaries, which she thinks may well be acceptable for an international audience, who just needs to have a general idea of some terms, but not acceptable for the members of those ethnicities, who want a more technical or specific definition of an ethnic term that they already know (cf. Murphy 1998: 13). Therefore, the dictionary's target audience needs to be taken into account when defining these terms and certainly much more is expected of a general-purpose dictionary which is to be used locally than of a dictionary intended for an international audience, especially a learners' dictionary. Also, if certain ethnicity and ethnicity-related terms are particularly sensitive in some societies, these considerations need to be taken into account.

Rader (1989) and Murphy (1998) both find that most of the definitions of ethnic terms in dictionaries are rather simplistic, being typically geographically based, although "this common failing has been criticised for over a century" (Rader 1989: 21). In her corpus, Murphy found just one dictionary attempting to go beyond a geographical definition, but this resulted in some problematic cultural stereotypes appearing in the work. Both Rader (1989) and Murphy (1998) find that some level of standardisation of ethnicity definitions in dictionaries is certainly necessary.

Despite claims of descriptiveness, dictionaries can have a prescriptive effect (Busse 2000: 166). Bearing in mind the authoritative role they have in the modern society, both in terms of perception and education (Veisbergs 2002) and even the legal role they play in some countries given that their definitions are used in court cases (Moon 1989), every effort should be taken to address the issues raised by the studies referred to above. Research has shown that modern dictionaries have greatly improved in this respect but that there still remains room for improvement, which is why these type of studies are very important.

2.2   Public reaction to the treatment of ethnic terms in dictionaries

Hauptfleisch (1993: 84-85) states that there are two possible directions in which a public may react to how some ethnic terms are treated in a dictionary. On the one hand, he explains that critical comments coming from individual users and reviewers are quite common and should not be "unduly worrying" as they may help improve the dictionary in its ensuing editions. On the other hand, a more serious threat, one which may affect not only the dictionary itself but also the lexicographer, in terms of his/her self-confidence and status, comes from pressure groups in a community. This may take the form of an organised protest against the definitions and labelling of some derogatory terms, and also against the very inclusion of some lexical items that have offensive connotations.

Hauptfleisch (1993) further gives a brief history of such public outcries. An often-cited essay is that of Burchfield (1980), who had edited the Oxford English Dictionary for thirty years, about his experiences following the publication of the definitions of some senses of the ethnic terms such as *Jew, Palestinian* and *Pakistani* in the *Concise Oxford Dictionary*. The pressure included boycotts of the dictionary and confiscations of its copies, until the contested definitions were amended. Even legal action was taken but the plaintiff lost the case. From such experiences, editors may learn that they need to devote maximum attention to sensitive terms, Burchfield (1980: 292) concludes and argues that dictionaries may aim to be normative only by the use of cautionary labels and/or symbols.

A judge also ruled in the favour of the publisher of the *Van Dale* dictionary, which included some negative expressions regarding Jews, as reported by Hauptfleisch (1993). As Burchfield, he also concludes that such situations make editors more aware of the sensitivity of some terms and more cautious in how they handle them, which is a positive thing. Still, Hauptfleisch advises against succumbing under pressure and keeping a cool head in adhering to professional standards in a dispassionate way.

3.   Ethnicity terms in various dictionaries

For the purposes of this paper, we inspected various online monolingual dictionaries and how they define ethnicities. We explore this issue using one specific ethnicity term only, but the findings are generalisable. We specifically sought entries for the ethnicity term *Albanian* as the definition for this entry proved to be contentious in the Montenegrin case study, which will be presented in section 4. We excluded the sense "Albanian language".

We used monolingual dictionaries which are freely accessible online, under the assumption that most users will first resort to these in an effort to look up an ethnicity term. As a result, different dictionaries are included in our study — most of them are general-purpose dictionaries, but there are also some learners' dictionaries. We covered dictionaries for several languages — English, as expected, had the largest number of free online monolingual dictionaries. We also inspected monolingual online dictionaries of German, Italian, Croatian, and Serbian, as these countries, amongst others, have considerable Albanian communities living in them. The number of the online monolingual dictionaries varied, depending on the language in question. We also included the Albanian free online monolingual dictionaries in the overview.

In the table below, we provide an overview of the dictionaries used and how they define the entry *Albanian* — both nouns and adjectives were inspected. Links to the relevant definitions of the entry *Albanian*, are given in the footnotes. The definitions from non-English dictionaries were literally translated into English.

Dictionary	Definition of *Albanian*
Merriam-Webster Dictionary[3]	n. a native or inhabitant of Albania adj. –
Cambridge Learner's Dictionary[4]	n. a person from Albania; adj. belonging to or relating to Albania, its people, or its language
Longman Dictionary of Contemporary English[5]	n. someone who comes from Albania adj. –
Oxford Advanced Learner's Dictionary[6]	n. (a person) from Albania adj. –
Online English Dictionary from Macmillan Education[7]	n. someone from Albania adj. relating to Albania, or its language or culture
Collins Advanced English Dictionary[8]	n. An Albanian is a person who comes from Albania. adj. belonging or relating to Albania, its people, language, or culture
Dictionary.com[9] *based on The Random House Unabridged Dictionary*	n. a native or inhabitant of Albania or Albany, N.Y. adj. pertaining to Albania, its inhabitants, or their language
The Free Dictionary (English)[10] based on *The American Heritage Dictionary of the English Language*, 5th Ed.	n. a. A native or inhabitant of Albania. b. A person of Albanian ancestry. adj. of or relating to Albania or its people, language, or culture
The Free Dictionary (German)[11] based on *Collins German Dictionary* — Complete and Unabridged 7th Ed.	n. someone with the citizenship of Albania adj. relating to Albania
Duden Wörterbuch (German)[12]	n. designation for an inhabitant of Albania adj. concerning Albania, the Albanians; originating from the Albanians, belonging to them
Digitales Wörterbuch der deutschen Sprache[13]	n. an inhabitant of Albania; someone with Albanian citizenship; someone who (originally) comes from Albania adj. –
Dizionario Italiano[14]	n. native or inhabitant of Albania adj. relating to Albania
Dizionario Italiano — Grandi Dizionari[15]	n. a native, an inhabitant of Albania; a person belonging to a minority speaking Albanian language in central Italy, Serbia, Macedonia, Greece adj. of Albania
Treccani Vocabolario[16]	n. an inhabitant of Albania. The name is extended also to the inhabitants of the Albanian colonies in Turkey, Greece, Dalmatia, and, in particular, those created by immigration in various regions of central Italy and Sicily. adj. of the Republic of Albania

Dizionario Internazionale[17]	Albanese — n. a native or an inhabitant of Albania adj. of Albania
Hrvatski jezični portal[18] (Croatian)	n. 1. a native or person holding citizenship of Albania; 2. a person that is Albanian by nationality. "Arbanas, Arnaut, Šiptar, Gegi" are listed as synonyms. adj. 1. relating to Albania or Albanians 2. undeveloped, obsolete
Online rečnik[19] (Serbian)	n. an inhabitant of Albania. "Arnaut, Šćipetar, Šiptar" are listed as synonyms. adj. –
Fjalor Shqip[20] (Albanian)	n. an indigenous resident of Albania or one of Albanian origin; a member of the Albanian nation adj. –
Fjalor i Gjuhës Shqipe[21] (Albanian)	n. a local resident of ethnic Albania or of ethnic Albanian origin; a member of the Albanian nation adj. –

Table 1:   Definition of the ethnic term *Albanian* in various monolingual online dictionaries

As can be seen, most of the dictionaries defined the noun using geographical criteria, defining the ethnicity term as "an inhabitant/a native of a COUNTRY", "someone from a COUNTRY". As pointed out by Rader (1989), such a designation may be problematic as there could be large communities living outside the borders of the country with whose name their ethnonym is related. We are purposefully not using the term their "homeland", as these could be autochthonous communities of people who do not hail from such a country but have always lived elsewhere. In this particular case, autochthonous communities of Albanians live in a number of countries outside Albania (Montenegro included). Also, there are large Albanian immigrant communities in many Western European and North American countries. In fact, more ethnic Albanians live outside Albania than inside it, which makes this kind of a definition even more problematic.

     Some dictionaries reduced the definition of the noun to just citizenship ("someone with the citizenship of a COUNTRY"). We also find this very problematic as citizenship may be held by members of ethnic minorities living in that particular country who do not belong under the related ethnic term, e.g. in this particular case, there are Serbs, Montenegrins and Greeks living in Albania, many of them probably holding the Albanian citizenship but, generally, they would not call themselves Albanians. Equating citizenship with how one identifies himself/herself may be the prevailing norm in some countries, but in the Balkans, for instance, this has never been the case. On the other hand, the definition is also problematic as there are many ethnic Albanians living, for instance, in Montenegro and having a Montenegrin citizenship, but some of

them would not identify themselves as Montenegrins.

As noted in the literature, complex issues are simplified in dictionaries and sometimes this is more justified if a dictionary is intended for an international audience, who just needs a general idea of what an ethnicity term that they are not familiar with stands for (Murphy 1998). Still, we must point out that even some international dictionaries did offer more apposite definitions when it came to the adjectival forms of the term, involving the issue of "culture" and including additionally the wording "relating to" in their definition, e.g. "belonging or relating to a COUNTRY, its people, language, or culture" (*Collins Advanced English Dictionary*) or "of or relating to a COUNTRY or its people, language, or culture" (*The Free Dictionary* based on *The American Heritage Dictionary of the English Language*). Namely, "belonging" and "pertaining" are too exclusive to accommodate for some of the cases discussed above (e.g. autochthonous communities of Albanians living outside Albania). "Relating to" is broader and more inclusive than these two forms, given that the ethnic communities living outside the country from whose name their ethnonym was derived, are always "related to" it in certain ways (culturally, language-wise, etc.). Going beyond geography to include culture (as well as the relation to the language), as typically one of the defining characteristics of an ethnicity, certainly is a step forward in defining ethnicity terms.

Some dictionaries from countries with large communities of Albanians offered more detailed descriptions. One such country is Italy, in which there are substantial Albanian immigrant communities. Of the four Italian dictionaries included in this review (these are all the online monolingual dictionaries for this language which we were able to find via Google search), two are quite simply based on geography, in the ways discussed above, and the other two invest more efforts in defining this community and use particularised definitions for this ethnicity. In *Dizionario Italiano — Grandi Dizionari*, this ethnicity also includes Albanian minorities living in several countries, who are thus defined through their language ("a minority speaking the Albanian language"), as well as geographically ("in central Italy, Serbia, Macedonia, Greece"), though this list of countries could also be contested by some as being too narrow. The *Treccani Vocabolario* includes immigrant communities in certain countries and regions, as well as colonies in three territories, but fails to include some significant autochthonous communities outside Albania (for instance, in Montenegro and North Macedonia). What is evident is that the lexicographers had in mind that the members of all these communities might be using their dictionary and a particularised definition would be more suitable in this case. Even though the said definitions could both be said to be lacking in some respects, this customised approach in defining an ethnicity whose members are amongst the dictionary's target audience is commendable in our opinion.

Croatia is a Balkan country with a small Albanian minority. The Croatian dictionary (*Hrvatski jezični portal*) defines this ethnic term in terms of geography and citizenship ("a native or person holding citizenship of Albania") and also as

follows: "a person that is Albanian by nationality", which could be character-ised as a circular definition. While a circular definition solves the problem of political correctness and inclusivity, it certainly is not precise enough from a lexicographic point of view. This dictionary entry also lists some other names for this ethnicity as synonyms, one of which Albanians now find politically incorrect or even offensive ("Šiptar"), with no label which would mark it as such.

The same is the case with another regional dictionary, a Serbian one (*Online rečnik*). Serbia has a substantial Albanian minority and one of its regions, Kosovo, inhabited mostly by Albanians, declared its independence (after a history of ethnic conflicts in the region), now recognised by a considerable part of the international community. This dictionary defines an Albanian in the simplest terms, as "an inhabitant of Albania", and lists the same term ("Šiptar") as a synonym without a label which would warn that this is a politically incor-rect or disparaging term. We find that not using labels to mark this ethnic term as offensive is not a good practice, especially bearing in mind that these are regional and local dictionaries, the target users of which include these minorities.

We also examined two Albanian online dictionaries. They both added the meaning of "a member of the Albanian nation", similarly to the second mean-ing in the Croatian entry, which, despite being inclusive, we commented on as being a circular definition.

In summary, most dictionaries used rudimentary definitions for ethnicities and in the literature we saw that this is recommended, at least to a certain level (Rader 1989; Murphy 1998). As suggested above, we cannot expect interna-tional dictionaries to have detailed definitions of all ethnicities, but we did see that some of them used more inclusive definitions, involving culture and the wording "relating to" (instead of exclusively "belonging to"), which we think could be an appropriate model for international dictionaries. Some dictionaries went beyond the standardised approach when treating a minority which is substantial in the region for which the dictionary is intended. This customised approach, although it was slightly flawed in some respects in the dictionaries we inspected, could be recommended for such cases, as it provides more inclu-sivity. In addition, if a certain ethnicity term could be seen as politically incor-rect or offensive, then it is certainly advisable to label it as such. Failing to do this might lead to a backlash from some communities, which could be justified, at least to some degree. We will describe one such extreme case in the next section.

4.   Problematisation of ethnicity and ethnicity-related terms in a dictionary: A case study from Montenegro

This section will focus on a case study from Montenegro. Namely, shortly after its publication, the first dictionary of the Montenegrin language provoked vio-lent reactions and protests from some Albanian parties and the Bosniak Party, representing two ethnic minorities living in Montenegro, due to its treatment of

certain ethnicity and ethnicity-related terms. The protests had an epilogue in the Parliament and eventually caused the publisher, the Montenegrin Association of Sciences and Arts (MASA), to revoke the dictionary.

The first volume of *Rječnik crnogorskog književnog i narodnog jezika* (English translation: *Montenegrin Dictionary of Vernacular and Literary Language*), which contained 12,018 words beginning with the letters A, B and V, was published in March 2016, as the most important project of Montenegrin lexicography, representing the first complete overview of the lexical complexity of the Montenegrin language and laying the foundations for the development of the dictionaries of this type in Montenegro. As stated in the Preface, p. XI, the dictionary itself, being general and descriptive, should reflect social, scientific and civilisational reality, and give the first complete presentation of the lexical structures of the Montenegrin language, of its functional and stylistic diversity; in addition to providing a linguistic contribution, it was also supposed to be of a great cultural and national identity significance.

Soon after its publication, the problem with the definitions of the term *Albanac* (Eng. translation: *Albanian*) and its derived forms *albanizacija* (Eng. trans.: *Albanisation*) and *albanizovati* (Eng. trans.: *to albanisate*), which were seen as offensive and wrong by the Albanian people living in Montenegro, grew into a big political issue and it was regarded as case of "culturocide". A group of intellectuals, linguists, political activists and representatives of different state and private institutions, including the Albanian and the Bosniak people living in Montenegro, harshly criticised the management of MASA and the authorial team of the dictionary. They argued that MASA did not have a very favourable opinion of co-life and multiculturalism in Montenegro, and that MASA had the agenda to define, shape, recommend, propagate and try to spread their own desires, frustrations, prejudices, stereotypes, fears, covert and overt hatred. An MP of Albanian ethnicity in protest even tore a few pages of the dictionary during a live session of the National Assembly. The specific problems raised are described below:

1. Firstly, a problem arose with the ethnic term *Albanac* (Eng. translation: *Albanian*), which is defined in this dictionary as "an inhabitant of Albania; someone who is originally from Albania" (p. 43)[22]. This definition was seen as too reductionist, given that a substantial autochthonous Albanian minority lives in Montenegro. It follows from the dictionary definition that every *Albanian* must be originally from Albania and that (s)he cannot be an autochthonous inhabitant of another country and be originally from it. The dictionary definition, the Albanian representatives argued, denied them their autochthonicity in Montenegro. The authorial team argued that this was a rudimentary definition, applied to every ethnic entry in the dictionary, but this did not appease the Albanian representatives, who accused MASA of having a hidden agenda of not representing them as autochthonous in Montenegro.

Therefore, applying the standardised approach (which was followed by

many international dictionaries, as we saw in Section 3), in this particular case was problematic, given the ethnic sensitivities in the country, some of them surrounding the issue of ethnic autochthonicity.

2. Another problem arose with the ethnicity-related terms *albanizacija* (Eng. trans.: *Albanisation*) and *albanizovati* (Eng. trans.: *to albanisate*), which were defined as "imposed" assimilation processes, implying aggressive actions, the critics argued. They argued that some other assimilation terms in this dictionary, did not suggest imposition — for instance, the critics referred to how *balkanizacija* (Eng. transl.: *Balkanisation*) and *amerikanizacija* (Eng. transl.: *Americanisation*) were defined in MASA's dictionary:

—   Balkanisation: "adoption of the characteristics of the languages and cultures of the Balkan peoples, adoption of the Balkan tradition" (p. 130);
—   Americanisation: "1. to give someone or to something the features of the American way of life and culture; 2. to receive American characteristics, way of life and thinking, to become similar to the Americans" (pp. 56-57).

In contrast, this is how *albanizacija* (Eng. trans.: *Albanisation*) was defined:

—   Albanisation: "1. to convert to Albanians; to impose the Albanian language, culture and customs on other peoples; 2. to became an Albanian; to get the characteristics of an Albanian" (p. 43).

As can be seen, the definition of the noun *balkanizacija* (Eng. transl.: *Balkanisation*) does not mention or imply any coercion, any imposition or any kind of oppression by either neighbouring or distant peoples. Moreover, *balkanisation* as here described happens naturally like acquiring a language. Similarly, *amerikanizacija* (Eng. transl.: *Americanisation*) indicates the introduction, acceptance, or receiving some of American characteristics, which seems to be carried out voluntarily by both those who provide the characteristics and those who accept them. Contrary to that, this dictionary suggests that Albanians imposed their culture, language and customs, and converted other nations into Albanians through the albanisation process. This was also reinforced in the example accompanying the entry, which also confirms the "imposition" implied in the definition:

> Asking himself, he also offered some answers — based on the experiences of Orthodox refugees from Albania who were exposed to systematic albanianisation, especially in the time after World War II war in 1945. (Zoran Lakić)

Semantically analysed, all these terms (*Albanisation, Americanisation* and *Balkanisation*) have the same semantic base resting on *assimilation* (linguistic, cultural or national), which MASA lexicographers defined without implying imposition (p. 93), as "the adjustment of the minority to the majority adopting the characteristics of the majority". However, they did not use this definition as

their standard for all assimilation processes, but defined them in a customised way, depending on the ethnicities involved. In this particular case, the lexicographers did not opt for a standardised approach when treating assimilation as an ethnicity-related term, which created problems.

In defining *albanizacija* (Eng. trans.: *Albanisation*), the MASA lexicographers were guided by their corpus, in which they opted to include various texts from the last 200 years. However, this means that the whole of the 19th century and the early 20th century were included, which are periods marked by a series of ethnic wars and political ideologies imbued with inter-ethnic hatred and intolerance, which the present-day society has been trying to overcome for decades now (to varying success). So, the problem with these ethnicity-related terms might be found in the corpus itself — as Wierzbicka argues (1995: 194), obsolete corpora do not reflect the contemporary reality. The lexicographers did not use labels to mark potentially disparaging meanings of the term or its use in a historical context.

3. Another problematised ethnicity term in the dictionary was "Agarjanin" (Eng. trans.: "Hagarian"[23]), defined as: "muslim, Turk; unbeliever, infidel" (p. 10) and accompanied by a corpus example illustrating the second meaning. The Bosniaks were especially critical of this definition as it implied an equation sign between Muslims, Turks, and infidels, they argued. The senses of the word were delimited with a comma and a semi-colon; numbers were not used to imply different meanings. No disparaging labels were used or notes on the term's historic use and meanings.

This particular case points to the importance of delimiting different meanings and marking those that are disparaging. This remains the issue, however, of what qualifies as "disparaging", as we have seen, different tendencies in the literature regarding such markings. In this case, however, a label marking the offensive uses and probably some note on the historical context of the word would have been needed.

The pressure from the Albanian and the Bosniak communities was such that the whole dictionary was revoked over a few definitions. As we have seen, in one case the problem was following a standardised definition of all ethnicity terms equally, which did not account for the autochthonous communities living outside the country from whose name their ethnonym was derived. In the second case, the lexicographers did not use a standardised approach to define assimilation processes relating to certain ethnicities, governed by their corpus which did not reflect a modern reality. In the third case, no disparaging labels were used to mark the offensive uses of the term, nor were there any delimitations between the non-offensive and the offensive meanings.

Following the outcry and criticism, the Parliament adopted a Resolution on the dictionary with recommendations to MASA to stop its distribution, which MASA did. A new, revised edition was to be issued, but this has not happened in the five years since the event. Perhaps such definitions would not

have caused an outcry in another country or, at least, the reaction would not have been equally harsh, but in a country like Montenegro, in which the issue of ethnicity is an extremely sensitive one, along with the issue of autochtonicity and religion (the lines along which many stark divisions are drawn in the society), these issues become a matter to which a lexicographer should devote maximum attention.

5.   Conclusions

In this paper we described the problems of defining ethnicity terms in dictionaries. We reviewed a number on online monolingual dictionaries and critically analysed their ethnic definitions, and we also described the case of *Rječnik crnogorskog književnog i narodnog jezika* (Eng. trans.: *Montenegrin Dictionary of Vernacular and Literary Language*), which was revoked over a few ethnicity and ethnicity-related terms.

We found that most of the ethnic definitions in dictionaries are rather simple, being typically geographically or citizenship-based, but some of the definitions also proved more inclusive — such was the case with the definitions involving culture, as well as with the adjectival forms which included the wording "relating to". Such definitions offered good models for international and learner's dictionaries. We also commented on the pros and cons of the customised approach to some ethnicity terms used in a small number of the dictionaries we examined. Having analysed the case of a few ethnicity and ethnicity-related terms in a Montenegrin dictionary and the public reaction which followed its publication, we concluded the same thing as Hauptfleisch (1993) and Burchfield (1980): editors and lexicographers should be very cautious in how they define and label ethnicity and ethnicity-related terms. As we have seen, sometimes rudimentary definitions will not suffice, while at other times lack of standardisation will create problems.

As suggested in the literature (Rader 1989; Murphy 1998), we argue that some level of standardisation of ethnicity definitions in dictionaries should certainly be established, but we also commend a partly customised approach when defining ethnicities which are amongst the target audience of a dictionary, particularly if the issue of ethnicity is a sensitive one in a particular society. It is the responsibility of lexicographers to examine more carefully the characteristics of such ethnic groups, their autochtonicity, culture and religion, to accompany the definitions related to ethnicity with non-ideologically imbued examples, and to mandatorily use appropriate labelling to mark the disparaging ethnicity terms and uses.

Endnotes

1.   https://dictionary.cambridge.org/dictionary/english/research
2.   This section is not retained in Landau (2001).

3.   https://www.merriam-webster.com/dictionary/Albanian
4.   https://dictionary.cambridge.org/dictionary/english/albanian
5.   https://www.ldoceonline.com/dictionary/albanian
6.   https://www.oxfordlearnersdictionaries.com/definition/english/albanian?q=albanian
7.   https://www.macmillandictionary.com/dictionary/british/albanian_1
8.   https://www.collinsdictionary.com/dictionary/english/albanian
9.   https://www.dictionary.com/browse/albanian?s=t
10.   https://www.thefreedictionary.com/albanian
11.   https://de.thefreedictionary.com/albaner
12   https://www.duden.de/rechtschreibung/Albaner
13.   https://www.dwds.de/wb/Albaner
14.   https://www.dizionario-italiano.it/dizionario-italiano.php?parola=albanese
15.   https://www.grandidizionari.it/Dizionario_Italiano/parola/A/albanese.aspx?query=albanese
16.   http://www.treccani.it/vocabolario/ricerca/albanese/
17.   https://dizionario.internazionale.it/parola/albanese
18.   http://hjp.znanje.hr/index.php?show=search
19.   https://onlinerecnik.com/leksikon/srpski/albanac
20.   http://www.fjalorshqip.com
21.   https://fjalorthi.com/shqiptar
22.   The definitions from the Dictionary given here were literally translated into English.
23.   Descendants from Abraham's son Ishmael, whose mother was Hagar.

References

Online dictionaries

Cambridge Learner's Dictionary. 2020. https://dictionary.cambridge.org
Collins Advanced English Dictionary. 2020. https://www.collinsdictionary.com/
Dictionary.com. The Random House Unabridged Dictionary 2020. https://www.dictionary.com
Digitales Wörterbuch der deutschen Sprache. 2020. https://www.dwds.de
Dizionario Internazionale. 2020. https://dizionario.internazionale.it/
Dizionario Italiano — Grandi Dizionari. 2020. https://www.grandidizionari.it/Dizionario_Italiano
Dizionario Italiano. 2020. https://www.dizionario-italiano.it/
Duden Wörterbuch. 2020. https://www.duden.de
Fjalor i Gjuhës Shqipe. 2020. https://fjalorthi.com
Fjalor Shqip. 2020. http://www.fjalorshqip.com
Hrvatski jezični portal. 2020. http://hjp.znanje.hr/
Longman Dictionary of Contemporary English. 2020. https://www.ldoceonline.com
Merriam-Webster.com. 2020. https://www.merriam-webster.com
Online English Dictionary from Macmillan Education. 2020. https://www.macmillandictionary.com
Online rečnik. 2020. https://onlinerecnik.com/leksikon
Oxford Advanced Learner's Dictionary. 2020. https://www.oxfordlearnersdictionaries.com
The Free Dictionary — The American Heritage Dictionary of the English Language. Fifth Edition. 2020.
     https://www.thefreedictionary.com

The Free Dictionary — Collins German Dictionary. Complete and Unabridged Seventh Edition. 2020. https://de.thefreedictionary.com

Treccani Vocabolario. 2020. http://www.treccani.it/vocabolario/

Other literature

Barnickel, K.-D. 1999. Political Correctness in Learners' Dictionaries. Herbst, T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary(?)*: 161-174. Tübingen: Niemeyer.

Brass, P.R. 1991. *Ethnicity and Nationalism: Theory and Comparison.* Newbury Park, CA: Sage.

Burchfield, R. 1980. Dictionaries and Ethnic Sensibilities. Michaels, L. and C. Ricks (Eds.). 1980. *The State of the Language*: 15-23. Berkeley: University of California Press.

Busse, U. 2000. Recent English Learners' Dictionaries and Their Treatment of Political Correctness. Mogensen, J.E., V.H. Pedersen and A. Zettersten (Eds.). 2000. *Symposium on Lexicography IX. Proceedings of the Ninth International Symposium on Lexicography, April 23–25, 1998 at the University of Copenhagen*: 165-201. Tübingen: Niemeyer.

Chandra, K. 2006. What is Ethnic Identity and Does It Matter? *Annual Review of Political Science* 9: 397-424.

Connor, W. 1984. Eco- or Ethno-nationalism? *Ethnic and Racial Studies* 7(3): 342-359.

Dieckmann, W. 1975. *Sprache in der Politik. Einführung in die Pragmatik und Semantik der politischen Sprache. Mit einem Literaturbericht zur 2. Auflage.* Heidelberg: Carl Winter.

Ezquerra, M.A. 1995. Political Considerations on Spanish Dictionaries. Kachru, B.B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*: 143-152. Tübingen: Niemeyer.

Francis, E.K. 1947. The Nature of the Ethnic Group. *American Journal of Sociology* 52(3): 393-400.

Gabbert, W. 2006. Concepts of Ethnicity. *Latin American and Caribbean Ethnic Studies* 1(1): 85-103.

Hauptfleisch, D.C. 1993. Racist Language in Society and in Dictionaries: A Pragmatic Perspective. *Lexikos* 3: 83-139.

Horowitz, D.L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Kalogjera, D. 2001. "Teške riječi" za leksikografa. *Filologija* 36–37: 263-271.

Landau, S. 1984. *Dictionaries. The Art and Craft of Lexicography*. New York: The Scribner Press.

Landau, S. 2001. *Dictionaries. The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.

Montenegrin Association of Sciences and Arts. 2016. *Rječnik crnogorskog književnog i narodnog jezika* (Montenegrin Dictionary of Vernacular and Literary Language). Podgorica: Montenegrin Association of Sciences and Arts. (revoked)

Moon, R. 1989. Objective or Objectionable: Ideological Aspects of Dictionaries. Knowles, M. and K. Malmkjær (Eds.). 1989. *ELR Journal* 3: 59-95.

Murphy, M.L. 1991. Defining Racial Labels: Problems and Promise in American Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 13(1): 43-64.

Murphy, M.L. 1998. Defining People: Race and Ethnicity in South African English Dictionaries. *International Journal of Lexicography* 11(1): 1-33.

Nissinen, S. 2015. *Insulting Nationality Words in Some British and American Dictionaries and in the BNC*. MA thesis. Tampere: University of Tampere.

Norri, J. 2000. Labelling of Derogatory Words in Some British and American Dictionaries. *International Journal of Lexicography* 13(2): 71-106.

Rader, J. 1989. *People and Language Names in Anglo-American Dictionaries. Dictionaries. Journal of the Dictionary Society of North America* 11: 125-138.

Rothschild, J. 1981. *Ethnopolitics: A Conceptual Framework*. New York: Columbia University Press.

Schutz, R. 2002. Indirect Offensive Language in Dictionaries. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 637-641. Copenhagen: Center for Sprogteknologi, Copenhagen University.

Ştefănescu, M. 2015. Dysphemisms for Ethnicity: Cautionary Labelling of Disparaging Ethnic Words in Some Romanian Dictionaries. *Studia Universitatis Babeş-Bolyai – Philologia* 60(2): 51-63.

Veisbergs, A. 2002. Defining Political Terms in Lexicography: Recent Past and Present. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 657-667. Copenhagen: Center for Sprogteknologi, Copenhagen University.

Wachal, R.S. 2002. Taboo or Not Taboo: That is the Question. *American Speech* 77(2): 195-206.

Wierzbicka, A. 1995. Dictionaries and Ideologies: Three Examples from Eastern Europe. Kachru, B.B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*: 181-196. Tübingen: Niemeyer.

Xu, Jieshun. 2002. Ethnic Group and Nationality: The Concepts and Their Relationships. *Ethnonational Studies* 1: 12-18.

Italian Dictionaries of Abbreviations and the Preparation of Entries of the Italian–Slovene Dictionary of Abbreviations

Mojca Kompara Lukančič, *Faculty of Criminal Justice and Security and Faculty of Tourism, University of Maribor, Slovenia (mojca.kompara@gmail.com) (mojca.kompara@um.si)*

Abstract: The present article addresses the two contemporary Italian dictionaries of abbreviations Malossini (1999) and Righini (2001) and outlines the position of the Italian abbreviations in some other Italian dictionaries, namely the monolingual Zingarelli (2000) and *Garzanti* (2021) and bilingual Slovene–Italian (Šlenc 2006) and Italian–Slovene (Šlenc 1997). The aim of the paper is to give an insight into the compilation of the abbreviation dictionary entries in monolingual Italian, bilingual in tandem with the Slovene language and specialised abbreviation dictionaries; highlight the elements present within the dictionary entries; present the deficiencies in compiling and examples of good practice; with the aim of introducing the outcomes in the compilation of the bilingual Italian–Slovene dictionary entries of the Slovene Contemporary Dictionary of Abbreviations. Within the paper we present the composition of the bilingual dictionary entries giving the reader an insight into the elements of the dictionary entry, namely the usage of language and field qualifiers, articulated prepositions, i.e. preposizioni articolate (composed of prepositions and definite articles), official translations and additional descriptions.

Keywords: ABBREVIATIONS, BILINGUAL DICTIONARIES, COMPILATION, DICTIONARIES, DICTIONARY ENTRY, EXPANSION, ITALIAN, LEXICOGRAPHY, SLOVENE, SPECIALISED DICTIONARIES

Opsomming: Italiaanse afkortingswoordeboeke en die voorbereiding van inskrywings van die Italiaans-Sloweense afkortingswoordeboek. In hierdie artikel word die twee hedendaagse Italiaanse afkortingswoordeboeke Malossini (1999) en Righini (2001) aangespreek en word die stand van Italiaanse afkortings in sommige ander Italiaanse woordeboeke, naamlik die eentalige Zingarelli (2000) en *Garzanti* (2021) en die tweetalige Sloweens–Italiaanse (Šlenc 2006) en Italiaans–Sloweense (Šlenc 1997) woordeboeke bespreek. Die doel van die artikel is om die samestelling van die afkortingswoordeboekinskrywings in eentalig Italiaanse, tweetalig Sloweens–Italiaanse/Italiaans–Sloweense en gespesialiseerde afkortingswoordeboeke weer te gee; om die elemente waaruit die woordeboekinskrywings bestaan, uit te lig; om die leemtes in die samestelling asook voorbeelde van goeie praktyk te beskryf; om sodoende die resultate in

die samestelling van die tweetalige Italiaans–Sloweense woordeboekinskrywings van die Slo-
weense Kontemporêre Afkortingswoordeboek toe te pas. In die artikel word die samestelling van
die tweetalige woordeboekinskrywings weergegee om die leser insig te gee in die elemente van die
woordeboekinskrywing, naamlik die gebruik van taal- en veldbepalers, geartikuleerde voorsetsels,
m.a.w. preposizioni articolate (saamgestel uit voorsetsels en bepaalde lidwoorde), amptelike verta-
lings en addisionele beskrywings.

Sleutelwoorde: AFKORTINGS, TWEETALIGE WOORDEBOEKE, SAMESTELLING,
WOORDEBOEKE, WOORDEBOEKINSKRYWING, UITBREIDING, ITALIAANS, LEKSIKOGRA-
FIE, SLOWEENS, GESPESIALISEERDE WOORDEBOEKE

Introduction

Abbreviations — grouped, joined, unified compositions of letters — often taken
from full words of phrases (Kompara 2018) present in almost any language
have the ability to materialise when needed i.e. **COVID-19** *(COrona VIrus Disease),*
SARS *(Severe acute respiratory syndrome),* etc. and vanish when not needed any-
more, i.e. in the case of company names such as **Beti** *(Belokranjska trikotažna
industrija),* or political parties such as **LDS** *(Liberalna demokracija Slovenije).* The issue
of the formation of new abbreviations in the Slovene language without being
properly normed has been thoroughly discussed by Kompara Lukančič (2018,
2017, 2009) Logar (2005) Verovnik and Logar (2006) and Verovnik (2018) and
Fajfar and Žagar Karer (2015). Given the absence of such a publication (Kompara
Lukančič 2018, 2017), we compiled a Slovene Contemporary Dictionary of Abbre-
viations which is currently in the reviewing process. In order to present the
process of creating such a dictionary, in the present paper we address the posi-
tion of abbreviations in the Italian dictionaries of abbreviations, monolingual
and bilingual dictionaries, which has led to the preparation of the Italian–Slovene
dictionary entries in the Slovene Contemporary Dictionary of Abbreviations. The
Italian–Slovene dictionary is a part of the Slovene Contemporary Dictionary of
Abbreviations which comprises over 20 languages, among them the Italian lan-
guage is on the third position, preceded by English and Slovene. Within the
paper we give an overview to the characteristics of the dictionary entries of the
Italian dictionaries of abbreviations from Malossini (1999) and Righini (2001)
and outline the position of the Italian abbreviations in some other Italian dic-
tionaries, namely the monolingual Zingarelli (2000) and *Garzanti* (2021) and
bilingual Slovene–Italian (Šlenc 2006) and Italian–Slovene (Šlenc 1997), and
present the Italian–Slovene dictionary entries in the Slovene Contemporary Dic-
tionary of Abbreviations.

Italian dictionaries of abbreviations

In Italian, dictionaries of abbreviations are not as common as in English (Kom-

para Lukančič 2020), but we can still rely on two works that give a relatively contemporary insight into abbreviations and were published in a rather similar time frame, Malossini in 1999 and Righini in 2001. Due to the fact that both dictionaries were published twenty years ago, we cannot expect to find up-to-date abbreviations, i.e. **COVID**, but they still give an insight into the hidden meaning, often not visible in the abbreviated letters, of abbreviations providing the reader with numerous expansions. In the Italian dictionaries of abbreviations, a random sampling method was used.

DidiSi — *Dizionario di Sigle Abbreviazioni e Simboli* **(Righini 2001)**

Dizionario di Sigle Abbreviazioni e Simboli — *DidiSi* (Righini 2001) is an Italian dictionary of abbreviations that includes over 10,000 alphabetically ordered Italian and foreign abbreviations in 14 different languages and over 30,000 expansions. Prior to the preface there is a list of 68 abbreviations used within the dictionary entries, denominating mainly language and field qualifiers, i.e. it. – italiano (*Italian*), chim. – chimica (*chemistry*). In the preface there is a short and very generic introduction to abbreviations, followed by a short introduction to the usage of the present dictionary. The dictionary is only available in paper format. Within the introduction on how to use the dictionary, the author mentions that abbreviations are not organised in the same way as normal language, as they follow their own rules of formation and consequently the way of reporting is, especially in the press, all uppercase or lowercase letters, with the occurrence of different layouts i.e. italics, bold, etc. Righini (2001) argues that in the dictionary entry the abbreviations are presented as uppercase, bold headwords. Foreign abbreviations are followed by the source languages, the field qualifiers and the translations into Italian, and some additional information when needed. The dictionary entries expand over 270 pages. Below the most representative general and specific characteristics of the dictionary entries, in alphabetic order, are explained.

Example 1

> **A** *1* altezza / area. *2* (*fis.*) ampere. *3* atomica (*bomba*). *4* amplificazione. *5* Austria (*anche targa autom.*). *6* nell'alpinismo, seguita da un numero, indica il grado di difficoltà della scalata artificiale. *7* ingl. Ace (asso, *nelle carte da poker e da ramino*). *8* (*biol.*) Adenina, *base purinica del DNA e RNA*. *9* Autostrada / *ted.* Autobahn / *fr.* Autoroute. *10* (*banca*) Azione. *11* (*anat.*) Arteria. *12* Alfiere (*nel gioco degli scacchi*). *13* (*ferr.*) Automotrice. *14* (*meteor.*) *centro di* Alta *Pressione*. *15* Alfa (A *nella compitazione internazionale*) / Ancona (A *nella compitazione italiana*).

Cf. the following examples 1, the entry is composed of an abbreviation written in bold, i.e. **A,** and followed by numerically ordered expansions or meanings. The Arabic numbers are written in bold and italics and are followed by expansion(s), which, in some synonymous cases, are divided by a slash, i.e. *9* Autostrada / *ted*. Autobahn / *fr*. Autoroute (*highway*). Every single expansion terminates with full stop. As seen from example 1, for some expansions, mainly the specialised ones, the field qualifier is also provided, i.e. *2* (*fis*.) ampere, where (*fis*.) denominates *physics*. The list of field qualifiers is given at the beginning of the dictionary, prior to the preface or introduction to the dictionary. With the dictionary entry the field qualifiers are given in brackets, usually clipped, in italics and positioned prior to the expansion, i.e. (*fis*.) *physics*, (*anat*.) *anatomy*. As seen from example 1 additional information is also present within the dictionary entry, i.e. *3* atomica (*bomba*) (*atomic bomb*) and it follows the expansion, in brackets and italics. The additional information is provided as a simple word (*bomba*) *bomb* or as a whole explanatory sentence (*anche targa autom*.) (*also car plate*), where the final word is clipped. In example 1 expansion 6, i.e. *6 nell'alpinismo, seguita da un numero, indica il grado di difficoltà della scalata artificiale* (*in mountaineering, followed by a number, it indicates the degree of difficulty of the artificial climb*) we notice only the explanatory sentence written in italics giving detailed information about the usage of the abbreviation. Another way of providing additional information is seen in expansion 8, i.e. *8* (*biol*.) Adenina, *base purinica del DNA e RNA* (*Adenine, purine base of DNA and RNA*), where the explanatory sentence is not in brackets but follows the expansion, is divided by a comma and is written in italics. We have to mention another way of providing additional information, namely expansion 14, i.e. *14* (*meteor*.) *centro di* Alta *Pressione* (*High Pressure center*), where we notice that the additional information is provided in italics within the expansion. In expansion 10, i.e. *10* (*banca*) Azione (*stock*), we notice some inconsistency in providing the additional information (*banca*) (*bank*). We do not know if it functions as a field qualifier, but it should be abbreviated and the word is not included in the list of abbreviations used in the dictionary, or as additional information, that according to expansion 3 should be placed after the expansion. Within the dictionary entry we also notice the introduction of language qualifiers, i.e. *7 ingl.* Ace (asso, *nelle carte da poker e da ramino*) (*ace, in poker and rummy cards*), which are positioned prior to the expansion, are abbreviated and written in italics, i.e. *ingl*. (*English*). In cases where the same abbreviation is used in more than one language the Italian expansion appears in the first position, followed by the expansions in foreign languages, i.e. *9* Autostrada / *ted*. Autobahn / *fr*. Autoroute (*highway*), as seen from the example, the expansions are divided by a dash and the alphabetic order of the foreign expansions is not preserved, as the German expansion is placed prior to the French.

Example 2

> **ACI** *1* Automobile club d'Italia. *2* Azione Cattolica Italiana. *3* Associazione Cartografica Internazionale. *4* Aero Club Italiano *o* d'Italia = AeCI. *5* Associazione Culturale Italiana / Alleanza Culturale Internazionale. *6* Aviazione Civile Italiana / Aviazione Civile Internazionale. *7* *ingl.* Airports Council International (Consiglio internazionale degli aeroporti). *8* Alleanza Cooperativa Internazionale.

Example 3

> **AeCI** Aero Club Italiano *o* d'Italia

As seen from example 2, within the dictionary entry in expansion 4, i.e. *4* Aero Club Italiano *o* d'Italia = **AeCI** (*Italian Aero Club or Aero Club from Italy*), there is the occurrence of two expansions of the same concept, namely *Aero Club Italiano* (*Italian Aero* Club) and *Aero Club d'Italia* (*Aero Club from Italy*) followed by the abbreviation that is also used in Italian for this concept after the symbol =, i.e. = **AeCI**. In example 3 we notice some inconsistency in providing the cross-reference, as the abbreviation **ACI** is not mentioned within the dictionary entry.

Example 4

> **ACNUR** Alto Commissario delle Nazioni Unite per i Rifugiati = UNHCR.

Within the dictionary entries we notice Italian and foreign abbreviations, in example 4 we notice the introduction of an Italian abbreviation, **ACNUR** that was originally English, i.e. **UNHCR**. As seen from the dictionary entry the expansion in Italian is followed by the symbol = and the original abbreviation **UNHCR**. A cross-reference is also provided as seen in example 5, where some inconsistency in spelling, namely the uppercase letters, i.e. Alto Commissario and Rifugiati (*High Commissioner and Refugees*) in example 4 vs the lowercase letters, i.e. alto commissariato and rifugiati (*High Commissariat and Refugees*) in example 5 as well as the occurrence of two different nouns, namely *Commissario* (*Commissioner*) and *commissariato* (*Commissariat*) is visible.

Example 5

> **UNHCR** *ingl.* United Nations High Commissioner for Refugees (alto commissariato delle Nazioni Unite per i rifugiati = ACNUR).

Some interesting example pairs of abbreviations

Below some particular example pairs of abbreviations are presented within joint examples from 6 up to example 10.

Example 6

> **A.C. di G.** Alta Corte di Giustizia
> **BdS** Banco di Sicilia

In example 6 there are two abbreviations composed of a preposition, namely *di* (*of, from*), that in the abbreviation **A.C. di G.** is fully spelt out, while in **BdS** it is abbreviated and forms a unified abbreviation.

Example 7

> **ACMONITAL** ACciaio MONetraio ITALiano.
> **ACOSER** Azienda COnsorziale SErvizi Reno (Bologna) *oggi* SEABO.
> **ACOTRAL** Azienda COnsortile TRAsporti Laziali *oggi* COTRAL.

In example 7 we encounter examples of acronyms, i.e. **ACMONITAL** ACciaio MONetraio ITALiano (*name of a company*), clipped to the first two or three letters. Such type of formation of abbreviations also used to be popular in Slovenia, i.e. **Beti**, **Nama**, etc. (Gložančev 1980). In example 7 we also notice the occurrence of acronyms that are outdated and at present a different abbreviation is used and introduced within the dictionary entry with the word *oggi*, (*today*), i.e. **ACOSER** Azienda COnsorziale SErvizi Reno (Bologna) *oggi* SEABO (*name of a company*), and **ACOTRAL** Azienda COnsortile TRAsporti Laziali *oggi* COTRAL (*name of a company*). Cross-references are ensured as both abbreviations, namely **COTRAL** and **SEABO**, are included as separate dictionary entries.

Example 8

> **ACL** (*elettron*.) *fr*. Affichage à Cristaux Liquides (visualizzazione a cristalli liquidi) = LCD.
> **LCD** (*elettron*.) *ingl*. Liquid Crystal Display (visualizzazione, schermo a cristalli liquidi).

Some inconsistency is seen in example 8 where for the French abbreviation **ACL**, composed of field and French language qualifier, followed by the French

expansion, the Italian explanation in brackets *(visualizzazione a cristalli liquidi)* (*liquid crystal display*), is provided, followed by a special symbol =, followed by the abbreviation **LCD**, that seems to be Italian, but is actually an English abbreviation. This is also seen in the cross-reference entry **LCD** where the entry is composed of an English language qualifier and consequently an English expansion. We also notice a difference in meaning provided in the Italian translation or explanation compared to **ACL**, namely *(visualizzazione, schermo a cristalli liquidi)*, (*display, liquid crystal screen*). The usage of the special symbol = is not explained in the preface. In examples 2 and 5 the = symbol is followed by the Italian equivalent of the abbreviation, but in examples 4 and 8, the same symbol is followed by the English equivalent.

Example 9

> **Bbk, BBk** *ted*. Deutsche Bundesbank (banca federale tedesca) = BUBA.
> **Bu.Ba., Buba** *1* (*pop*.) *ted*. Deutsche Bundesbank (Banca federale tedesca = Bbk).
>   *2* Bundesbahn (ferrovie federali tedesche).

In example 9 we notice some inconsistency in the two dictionary entries that should function as cross-references. The dictionary entry **Bbk, BBk** is followed by a language qualifier, the German expansion and in brackets the Italian explanation, followed by the special symbol = and the equivalent **BUBA**. Within the dictionary we do not find the dictionary entry for the abbreviation **BUBA**, but we encounter a differently spelt dictionary entry, namely **Bu.Ba., Buba** that has two expansions. Within the first expansion we notice the same abbreviation meaning. The entry is composed of the usage of a field qualifier (*pop*.) denominating *popolare* or *popolarmente* (*popular or popularly*), that is not included in the entry **Bbk**, **BBk**, followed by the language qualifier *ted*., the German expansion and the Italian explanation or translation in brackets *(Banca federale tedesca = Bbk)* (*German federal bank*), that is different from the explanation in the expansion of the entry **Bbk**, **BBk**. The difference is seen in the usage of a capitalised noun, *banca* vs *Banca*, and the occurrence of the equivalent **Bbk** in brackets in the dictionary entry **Bu.Ba., Buba**. The abbreviations are not systematically presented, in the entry **Bbk**, **BBk**, where the headword functions as a synonym the equivalent is **BUBA** and not **Buba** as suggested by the cross-reference. In the cross-reference entry **Bu.Ba., Buba**, where the headword functions as a synonym, after the Italian explanation and symbol there is only the abbreviation **Bbk** mentioned. Also, in the entry **Bbk**, **BBk** the equivalent abbreviation **BUBA** is not within brackets, in the entry **Bu.Ba., Buba** the equivalent **Bbk** is within brackets.

Example 10

> **BO** *1* Bologna. *2 ingl.* Box Office (botteghino di cinema o teatro). *3 ingl.* Body Odour (odore del corpo, *nel senso di odore sgradevole*). *4 sp.* Boletín Oficial (Gazzetta ufficiale della Spagna) = BOE.
>
> **BOE** *1 ing.* Barrel Oil Equivalent (equivalente di un barile di petrolio, *unità di riferimento per altri combustibili*). *2 sp.* Boletín Oficial del Estado (gazzetta ufficiale dello Stato).

Some inconsistency is also seen in example 10 where in the dictionary entry **BO**, i.e. *4 sp.* Boletín Oficial (Gazzetta ufficiale della Spagna) = BOE (*Spanish official gazette*) after the language qualifier and expansion, in brackets we notice the explanation or translation, namely *Gazzetta ufficiale della Spagna* (*Spanish official gazette*) that is followed by symbol = and the equivalent abbreviation **BOE**. In the cross-reference dictionary entry **BOE** in the second expansion we notice, i.e. *2 sp.* Boletín Oficial del Estado (gazzetta ufficiale dello Stato) (*Official gazette of the state*) that is composed of language qualifier, followed by the expansion that is slightly different from the one in **BO**, but preserves the same meaning, and in brackets the explanation or translation, namely *gazzetta ufficiale dello Stato* (*Official gazette of the state*) where we notice that the word *gazzetta* (*gazette*) is not capitalised, but the word *Stato* (*state*) is. It is not clear why the explanation or transition is provided in such way. We also notice the absence of the equivalent abbreviation **BO** in the entry **BOE**. Consequently, cross-referencing is not adequately presented.

Below, as exceptions, the explanations of two culturally and historically specific abbreviations are presented.

Example 11

> **EUR** *1* Esposizione Universale di Roma (*prevista nel 1942, non tenuta a causa della guerra*). *2 oggi* il quartiere residenziale sorto in quella zona. 3 *abbreviazione ufficiale della moneta euro* (dal 1° gennaio 1999 in tutti gli Stati partecipanti), *obbligatoriamente tutta maiuscola.*

In example 11 we notice the occurrence of historically labelled abbreviations, namely in the first expansion, i.e. *1* Esposizione Universale di Roma (*prevista nel 1942, non tenuta a causa della guerra*) (*Universal Exposition of Rome, scheduled for 1942, not held due to war*), where an explanatory sentence is provided stating that the event was never performed, but the abbreviation was still used enough to be included in the present dictionary. Within the second expansion, i.e. *2 oggi* il quartiere residenziale sorto in quella zona (*today the residential district built in that area*), we notice the time qualifier *oggi* (*today*) and an explanatory sentence

stating that the abbreviation **EUR** meant the *Universal Exhibition in Rome*, *Esposizione Universale di Roma* but was transferred to the present residential area of the aforementioned exhibition. The expansion denoting the currency, in use in Italy since 1999, is only placed as the 3rd expansion, *3 abbreviazione ufficiale della moneta euro* (dal 1° gennaio 1999 in tutti gli Stati partecipanti), *obbligatoriamente tutta maiuscola* (*official abbreviation of the euro currency, since 1 January 1999 in all participating countries, compulsory all upper case letters*). We notice that the number is not written in bold and we also notice a long explanatory sentence providing information about the official abbreviation of the euro currency, followed in brackets by some additional information stating that the currency was adopted by all participating countries on the 1st January 1999 and followed by an orthographic remark, i.e. *obbligatoriamente tutta maiuscola* (*compulsory all upper case letters*) stating that the abbreviation is obligatorily written in upper case letters.

Example 12

> **FNAC** *fr*. Fédération Nationale d'Achats des Cadres [Federazione nazionale d'acquisto dei quadri (*quadri in senso aziendale*), *iniziale gruppo d'acquisto divenuto grande catena internazionale di negozi specializzati nella vendita al miglior prezzo di prodotti tecnologici: dischi, libri, fotografia, hi-fi, TV, ecc. e in prodotti culturali: mostre, incontri e dibattiti, fondata del 1954 a Parigi*].

Example 12 is interesting as it was originally a French abbreviation. The dictionary entry is composed of the entry word in bold, followed by the language qualifier, *fr*. and the French expansion. We notice two types of brackets; the dictionary entry starts with square brackets where a detailed explanation in Italian, functioning as encyclopaedic data of the term, is given. In the first part an Italian equivalent of the expansion is provided, namely *Federazione nazionale d'acquisto dei quadri* (*National purchasing federation of tables*), followed by an additional explanation of the Italian term *quadri*, namely (*quadri in senso aziendale*) (*tables in the economic sense*) in brackets and in italics, and followed by the longer explanation of the abbreviation **FNAC** provided in italics, *iniziale gruppo d'acquisto divenuto grande catena internazionale di negozi specializzati nella vendita al miglior prezzo di prodotti tecnologici: dischi, libri, fotografia, hi-fi, TV, ecc. e in prodotti culturali: mostre, incontri e dibattiti, fondata del 1954 a Parigi* (*initial purchasing group that has become a large international chain of stores specialising in the sale of technological products at the best price: records, books, photos, hi-fi, TV, etc. and in cultural products: exhibitions, meetings and debates, founded in 1954 in Paris*).

Examining Righini's work we have to say that it is a great and fairly systematically compiled work. In the dictionary the language and field qualifiers are present when needed and systematically, foreign expansions are given after the

head word and are followed by the explanation or translation, provided in Italian, the latter is provided in brackets and in italics, additional information is also provided in brackets (also square brackets) and in italics. In the dictionary we encountered some compilation issues that are provided above. Translations and additional information or encyclopaedic data, as well as cross-references and in some cases field qualifiers, are among the linguistic issues that cause confusion.

Dizionario delle Sigle e degli Acronimi (Malossini 1999)

Dizionario delle Sigle e degli Acronimi (Malossini 1999) is an Italian dictionary of abbreviations that has over 8,000 Italian and foreign abbreviations and over 11,000 expansions. The dictionary extends to over 300 pages. In the short preface the author explains that the abbreviations predominately come from the field of informatics, telecommunication, institutions, associations, economy and transport. Prior to the preface there is no list of language and field qualifiers provided, the main reason being the fact that such elements are not included in the dictionary entries even if the dictionary does include foreign abbreviations in addition to Italian, namely well-known examples from English, i.e. BBC; German, i.e. ABS; Latin, i.e. a.a. etc. As seen from the dictionary entries, the foreign abbreviations are not translated, and no additional explanation or description is provided. The composition of the dictionary entry is simple, the abbreviation provided in bold is followed by a list of alphabetically ordered expansions, which are separated by a black dot. Here and there, within the dictionary entries, we notice in brackets the occurrence of additional information provided as simple words or short patterns.

Example 13

> **A** Airbus Industrie • Alfiere (scacchi) • Alta pressione (meteoroliga) • Area (geometria) • Assicurata (poste) • Asso (carte da gioco) • Associazione • Atomica (bomba) • Atto (teatro) • Australe • Austria • Automotrice • Autore • Autostrada • vitamina

In example 13 we notice a typical dictionary entry, composed of alphabetically ordered expansions, all written with a first capital letter, except for the last expansion, i.e. *vitamina* (*vitamin*), as it represents a set phrase. The expansions are separated by a black dot and in some cases there is additional information provided in brackets, usually a word or phrase that gives the reader the basic information of the field of usage of the abbreviation, i.e. Area (geometria) • Assicurata (poste) • Asso (carte da gioco).

Example 14

> **AC** Access Control • Aeroclub • Air Canada (cod. IATA) • Air Conditioning • Alter-
> nating Current • Assegno circolare • Automatic Computer • Aviazione civile • Azione
> cattolica

As seen in example 14 there are also foreign expansions present within the dic-
tionary entry, for English, i.e. Air Conditioning, Alternating Current, etc. and
for Italian, i.e. Aviazione civile (*Civil Aviation*), Azione Cattolica (*Catholic
Action*), but there is no information concerning the language qualifier prior to
the expansion and also no translation of the foreign expansions into Italian and
no explanations. Those were all elements that we encountered in Righini's dic-
tionary (2001). In Malossini's (1999) work we notice the introduction of addi-
tional information in brackets within the expansion *Air Canada (cod. IATA)*,
stating that it is an international code.

Example 15

> **af** Afganistan (cod. Internet)

In example 15 we notice the appearance of additional information in brackets,
i.e. (cod. Internet), *codice Internet*, in English *internet code*.

Example 16

> **BBPR** Banfi, Barbiano, Peressutti e Rogers (studio di architetti e urbanisti)

Example 16 draws our attention as we notice that the expansion is composed of
surnames and in brackets there is an explanation *(studio di architetti e urbanisti)*
(office/group of architects and urban planners) that functions as additional infor-
mation, translated into English it is (*a group of architects and urbanists*).

Example 17

> **BBS** Bulletin Board System [Service]
> **bcc** blind courtesy [carbon] copy

As seen in example 17 within the entries we also notice some cases where
square brackets are used, i.e. [Service] and [carbon]. The function of the square
brackets is to give the reader the information that both words are used, namely
system and *service* within the expansion of the abbreviation **BBS** and *carbon* and
courtesy within the expansion of the abbreviation **bcc**.

Example 18

> **CEKA** C'resvyciàjnaja komissija (polizia di Stato sovietica)

In example 18 we encounter a Russian abbreviation. The expansion is provided in Russian and is followed by an explanation in brackets (*polizia di Stato sovietica)* stating that it is the denomination of (*militia in the Soviet Union*).

Malossini's *Dizionario delle Sigle e degli Acronimi* (1999) was the first contemporary dictionary of Italian abbreviations, composed of a simple entry structure where the reader misses apart from the language and field qualifiers mainly the translations of foreign abbreviations. Just two years after Malossini's publication Righini's (2001) *Dizionario di Sigle Abbreviazioni e Simboli — DidiSi* was published, having a more detailed entry structure being composed of language and field qualifiers, translations and additional information. Being published twenty years ago both works needs to be improved and updated in order to cope with the current abbreviations. A digitalized version, one that could be periodically updated, of both dictionaries is also seen as a positive improvement on both works.

Other Italian dictionaries of abbreviations

Among other dictionaries of Italian abbreviations, it is worth mentioning the *Dizionario di abbreviature latine e italiane* (Cappelli 1899) which is more of a diachronic dictionary meant for individuals who are interested in readings and transcriptions of medieval texts and documents, both in Latin and Vulgar Italian, because in the two languages the system of abbreviations is widely used due to the complexity of the medieval system of abbreviations. The dictionary extends to 673 pages, starts with a long preface and is followed by the dictionary of Latin and Italian abbreviations.

     We encounter other abbreviations in paper monolingual Italian dictionaries, included in a more or less extensive appendix at the beginning or the end of a dictionary, as is the case of *Zingarelli* (Zingarelli 2000), where we encounter 23 pages of abbreviations in the appendix at the end of the dictionary. There are approximately 4,000 dictionary articles. Some examples are presented below.

Example 19

> **CAF**
> *1* Commissione d'Appello Federale (gioco del calcio) *2* (*comm.*) *fr*. Coût, Assurance, Fret (costo, assicurazione, spesa) = CIF, CAN *3* (chim.) Cloranfenicolo *4* Craxi Andreotti Forlani (alleanza politica tra DC e PSI della fine degli anni '80) *5 fr*. Club Alpin Français (Club alpino francese)

As seen from example 19 the dictionary entry is composed of several expansions that do not follow the alphabetical order of inclusion. Expansions are introduced with Arabic numbers written in bold and italics. As seen from the dictionary entry the field and language qualifiers are included after the Arabic number and prior to the expansions, both in italics and abbreviated, field qualifiers are in brackets. As seen from example 19 foreign abbreviations are included within the dictionary entry, i.e. Coût, Assurance, Fret (costo, assicurazione, spesa), (*cost, insurance, expense*), after the foreign expansion there is the Italian translation present in brackets, followed by the synonymous foreign abbreviation after the symbol =, i.e. **CIF**, **CAN**. Some additional information is also present within the 4th expansions in brackets, i.e. (alleanza politica tra DC e PSI della fine degli anni '80) (*political alliance between DC and PSI in the late 1980s*), giving additional meaning to the expansion of the abbreviations composed entirely of surnames. In the monolingual online dictionary *Garzanti* (2021), Italian and foreign abbreviations are included among the dictionary entries, as seen from example 20.

Example 20

BBC
  sigla
  ingl.: *British Broadcasting Corporation* (Ente britannico di radiodiffusione).
FIAT
  sigla
  Fabbrica Italiana Automobili Torino.

As seen from example 20 the dictionary entry is simple, composed of the headword written in bold and followed by the denomination *sigla*, in English *abbreviation*.[1] Within the dictionary entry the language qualifiers are included, i.e. *ingl.*, *inglese* (*English*) abbreviated and written in italics and followed by colon and the expansion. Foreign expansions are written in italics and followed by the explanation or translation in brackets, Italian expansions are not written in italics and there is no description or additional information given.

Within the bilingual encoding dictionaries, it is worth mentioning the *Veliki slovensko–italijanski slovar* (Šlenc 2006) where we do not encounter abbreviations, except for lexical and linguistic ones, i.e. ital., (*Italian*). More abbreviations are encountered in the bilingual decoding dictionary *Veliki italijansko–slovenski slovar* (Šlenc 1997) where in the appendix we find a list of approximately 1,000 abbreviations. The dictionary entry is simple, as seen in example 21 the expansions written in bold are alphabetically ordered and introduced by Arabic number in bold and followed by the Slovene translations. In the dictionary entry **AA** we notice the occurrence of abbreviated field qualifiers, i.e. MAT. (*mathematics*) GEOGR. (*geography*) prior to the expansion. In the dictionary

entry **A.C.I.** we also notice the occurrence of gender denomination prior to the expansions, i.e. *f*, for *femminile*, (*feminine*) and *m*, for *maschile*, (*masculine*).

Example 21

A
   1 anno leto **2 ara** ar
AA
   1 MAT. **altezza** višina **2** GEOGR. **Alto Adige** Južna Tirolska
A.C.I.
   1 *m* **Automobile Club d'Italia** Italijanska avtomoto zveza **2** *f* **Azione Cattolica Italiana** Italijanska katoliška akcija

---

The Italian–Slovene dictionary entries of the Slovene Contemporary Dictionary of Abbreviations

The aim of the paper is to observe how are abbreviations included into the Italian dictionaries of abbreviations, monolingual dictionaries and in bilingual dictionaries in tandem with Slovene, and use the examples of good practice in the compilation of the Slovene Contemporary Dictionary of Abbreviations (Kompara Lukančič 2017) that has been a topic of lexicographic interest in Slovenia since the first attempts from 2006 and 2011 were prepared, namely the *Slovarček krajšav* (2006) and the *Slovar krajšav* (2011).

The Slovene Contemporary Dictionary of Abbreviations is an alphabetically ordered dictionary that comprises of approximately 10,000 dictionary entries and over 30,000 expansions (Kompara Lukančič 2017) and is based on the two dictionary attempts from 2006 and 2011. The dictionary was also compiled thanks to the development of the algorithm for automatic recognition of abbreviations and abbreviation's expansions in electronic texts (Kompara Lukančič 2011), followed by the transfer of all dictionary entries to the *Termania* dictionary mask, and later transferred to two word documents, the first comprising of abbreviations of Slovene origin and the second comprising of foreign abbreviations, i.e. Italian, Latin, Spanish, English, etc. The first data base, namely the one comprising of Slovene abbreviations is currently in the revision process and we estimate it to be ready for publication in late 2021.

Within the compilation of the dictionary some steps were automatised, some, mainly due to its complexity, were manual. Due to its complexity, field qualifiers were manually added to the expansions. The editors, who are specialists in dictionary editing and compilation, prepared a list of abbreviated field qualifiers following steps from bilingual dictionaries (also Šlenc 1997, 2006) and determined the rules when an expansion needs the field qualifier. The entirely automatised step in compilation is found in language identification where the Presis analyser was used (Kompara Lukančič and Holozan 2011).

The analyser functions in the following way: in the first stage the expansion is sent to the Slovene analyser. If the analyser is successful, the language code "sl" is assigned; if not the language code "sl-x" is assigned (Slovene — to be manually checked). In cases where there is no code the Italian analyser is applied and if it is successful, the language code "it" is assigned, if not the language code "it-x" is assigned (Italian — to be manually checked). The same procedure is applied for all the languages included in the dictionary as a whole (Kompara Lukančič and Holozan 2011). The Italian–Slovene dictionary entries are part of the second data base, namely the one composed of foreign abbreviations. The extraction process was manual and automatic. For the automatic extraction of the Italian abbreviations, the online newspaper Delo was used. Monolingual, bilingual, specialised dictionaries and encyclopaedias were used for the manual extraction of the Italian abbreviations. The abbreviations were included in the dictionary if they matched with the expansions and occurred at least twice.

In presenting the Italian–Slovene dictionary entries of the Slovene Contemporary Dictionary of Abbreviations a random sampling method was used. The overall characteristic of all dictionary entries is the inclusion within the entries of abbreviations with more than one expansion, as seen in examples 22 and 28 where we encounter more than one expansion of the abbreviation. Following Righini's (2001) structure in adding translations, we decided that all Italian expansions are followed by the official translations or description and in several cases additional information is also provided, i.e. in examples 23 and 24. The official translations are set translations retrieved from dictionaries, encyclopaedias, databases, web pages, the Slovenian orthography, and the Official Gazette.

As seen in the examples below from 22 to 28, within the dictionary entries we notice the abbreviated language qualifiers, *it.*, (*Italian*). The language qualifier *it.* is present within the dictionary entries due to the fact that within the dictionary part comprising of foreign abbreviations the expansions of the abbreviations are in different languages, namely for **AA** there are English and Latin expansions present within the foreign language abbreviations, but for the purpose of the present paper only the Italian ones are presented. Below a detailed description of the examples of the Italian–Slovene dictionary entries is presented.

Example 22

AA
   it.: altezza ► višina
   it.: Alto Adige (geo.) ► Južna Tirolska

As seen in example 22 the dictionary entry is composed of two expansions, prior to the expansion there is the abbreviated language qualifier *it.*, followed by a

colon and the Italian expansion, when necessary the field qualifiers, i.e. (geo.), (*geography*) are included in brackets, following the mark ► there is the translation or equivalent in Slovene. A list of field and language qualifiers is provided prior to the preface of the dictionary.

Example 23

> **aff.**
>   it.: affezionatissimo ► vdani ◊ v pismih

In cases as demonstrated in example 23, the Italian expansion *affezionatissimo*, in English *very fond*, is followed by the Slovenian translation *vdani*, in English *devoted*, and after the symbol ◊ also an explanatory sentence, *v pismih*, in English *in letters*, which explains the usage of the abbreviation in correspondence.

Example 24

> **Agip**
>   it.: Azienda Generale Italiana Petroli (ekon.) ◊ italijanska naftna družba
> **ALITALIA**
>   it.: Aerolinee Italiane Internazionali (let.) ◊ italijanska državna letalska družba

In the dictionary entries from example 24, the expansions do not have an official translation, because the original abbreviations are proper names or names of a company, i.e. Agip, ALITALIA. To distinguish such cases from cases where we have official translations, we introduce a short and concise explanatory sentence after the symbol ◊.

Example 25

> **alla ted.**
>   it.: alla tedesca ► po nemško

Among the Italian abbreviations there are also cases of joint prepositions and definite articles, namely the *preposizioni articolate*, articulated prepositions that are composed of prepositions combined with a definite article, i.e. alla ted. Some similar examples from Righini's dictionary were presented above in example 6. We wanted to highlight such cases providing an example, i.e. alla ted., being aware that such examples are limited.

Example 26

> **alleg.**
>   it.: allegato; tudi all. ▸ priložen, priloga ○ pril.
> **all.**
>   it.: allegato; tudi alleg. ▸ priložen, priloga ○ pril.

In example 26 we encounter synonymity of the two abbreviations, namely **alleg.** and **all.** In both dictionary entries the headword is followed by the abbreviated language qualifier *it.*, colon and the Italian expansion, after semi colon and the word *tudi*, in English *also*, the synonymic abbreviation appears, i.e. all. and alleg., followed by the symbol used for translations ▸ and the Slovene translations, in this case we have two possible translations, *priložen* and *priloga*, in English *attached* and *attachment*. At the end of the entry there is also the Slovene equivalent abbreviation present, i.e. pril. introduced with the symbol ○. As seen in example 26 the cross-reference of the foreign abbreviation is provided. In the Slovene part of the dictionary the cross-reference for **pril.** is also provided.

Example 27

> **B.U.**
>   it.: Bollettino Ufficiale (prav.) ▸ Uradni list ○ Ur. l., Urad. l., Urad. list, Ur. list, UL

In example 27 there is more than one Slovene equivalent abbreviation present, in such cases all cross-references are included as separate dictionary entries. The entry is simple, composed of language qualifier, followed by the Italian expansion and field qualifier in brackets, the Slovene translation is provided after the symbol ▸ and the Slovene equivalent abbreviations after the symbol ○.

Example 28

> **C.C.**
>   it.: Codice Civile (prav.) ▸ Civilni zakonik ○ CZ
>   it.: Corte Costituzionale (prav.) ▸ Ustavno sodišče ○ US
>   it.: Corte dei conti (prav.) ▸ Računsko sodišče ○ RS
>   it.: Corte di Cassazione (prav.) ▸ Kasacijsko sodišče ○ KS

In example 28 we notice the occurrence of a series of Italian alphabetically ordered expansions that are accompanied with the abbreviated Slovene field qualifiers in brackets, i.e. (prav.), *pravo* in English *law*. The list of field qualifiers is provided prior to the dictionary prefaces. Due to the fact that all concepts of

the Italian expansion also exist in Slovene, the translations are introduced with the symbol ► and followed by the official Slovene translations. Following the translations and introduced by the symbol ○ are the Slovene equivalents of the abbreviations, namely **CZ**, **US**, **RS**, and **KS**. Within the Slovene part of the dictionary the cross-references of the previously mentioned Slovene abbreviations are provided among the dictionary entries.

Conclusion

Abbreviations are part of our everyday reality and are present in our lives and languages. They emerge out of current situations, i.e. COVID, relatively quickly on a daily basis. Some remain in a language forever; some disappear or are forgotten after the situation or concept is not an issue anymore. This fast occurrence and development of abbreviations calls to mind the necessity of up-to-date and functional abbreviation dictionaries. Unfortunately, as seen from the presented Italian dictionaries of abbreviations, too often the publication of a dictionary or a new edition of the already existing one does not occur in twenty years, as it is the case with the two Italian abbreviation's dictionaries (Righini 2001, Malossini 1999). The two analysed paper dictionaries differ in the number of entries included and characteristics of entries. We encounter inconsistencies among the inclusion and exclusion of translations, the language and field qualifiers, and additional information. In reference to the monolingual and bilingual dictionaries presented, a common characteristic is the fact that they are of limited entry number, because the abbreviations in such dictionaries are only included within the appendix and as such treated marginally. The overview of the position of abbreviations in all three types of dictionaries helped us prepare an adequate structure of the Italian–Slovene dictionary entries that are part of the Slovene Contemporary Dictionary of Abbreviations that is in preparation. Functional, contemporary, structurally adequate dictionaries of abbreviations are the result of combining the past and present concepts of lexicography.

Endnote

1.   Denominations and consequently translations may differ from author to author, i.e. *sigla* can also be translated as *acronym*.

References

Cappelli, A. 1899. *Dizionario di abbreviature latine ed italiane.* https://archiviodistatotorino.beniculturali. it/strumenti/dizionario-abbreviature-italiane/
      Date of Access: 5 February 2021.
Fajfar, T. and M. Žagar Karer. 2015. Strokovnjaki in drugi uporabniki terminologije kot naslovniki terminoloških slovarjev. *Jezikoslovni zapiski* 21(1): 7-21.

Garzanti: https://www.garzantilinguistica.it/ricerca/?q=dizionario
  Date of Access: 5 February 2021.

Gložančev, A. 1980. Kratice v imenih delovnih organizacij. *Jezik in slovstvo* 26(2): 75-78.

Kompara Lukančič, M. 2006. *Slovarček krajšav*. Ljubljana: Inštitut za slovenski jezik Fran Ramovš ZRC SAZU.

Kompara Lukančič, M. 2009. Prepoznavanje krajšav v besedilih. *Jezikoslovni zapiski* 15(1–2): 95-112.

Kompara Lukančič, M. 2011. *Slovar krajšav*. Kamnik: Amebis, Termania.

Kompara Lukančič, M. 2017. Zasnova novega slovarja krajšav. *Jezikoslovni zapiski* 23(1): 77-92.

Kompara Lukančič, M. 2018. *Sinhrono-diahroni pregled krajšav v slovenskem prostoru in sestava slovarja krajšav.* Maribor. Univerzitetna založba Univerze.

Kompara Lukančič, M. 2020. An Overview of English Dictionaries of Abbreviations. *Lexikos* 30: 171-196.

Kompara Lukančič, M. and P. Holozan. 2011. What is Needed for Automatic Production of Simple and Complex Dictionary Entries in the First Slovene Online Dictionary of Abbreviations Using Termania Website. Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, Slovenia, 10–12 November 2011*: 140-146. Ljubljana: Trojina, Institute for Applied Slovene Studies.
  https://elex2011. trojina.si/Vsebine/proceedings/eLex2011-17.pdf
  Date of Access: 5 February 2021.

Logar, N. 2005. Norma v slovarju sodobne slovenščine: zloženke in kratice. *Družboslovne razprave* 21(48): 211-225.

Malossini, A. 1999. *Dizionario delle Sigle e degli Acronimi*. Milano: A. Vallardi.

Righini, E. 2001. *Dizionario di sigle abbreviazioni e simboli.* Bologna: Zanichelli.

Šlenc, S. 1997. *Veliki italijansko–slovenski slovar.* Ljubljana: DZS.

Šlenc, S. 2006. *Veliki slovensko–italijanski slovar.* Ljubljana: DZS.

Termania: http://www.termania.net/
  Date of Access: 5 February 2021.

Verovnik, T. 2018. Obravnava kratic v prenovljenih pravopisnih pravilih. *Jezikoslovni zapiski* 24(2): 43-54.

Verovnik, T. and N. Logar. 2006. O jeziku, stilu i utjecaju slovenskih tiskanih oglasa. Granić, J. (Ed.). 2006. *Jezik i mediji: jedan jezik: više svjetova:* 743-752. Zagreb, Split: Hrvatsko društvo za primijenjenu lingvistiku.

Zingarelli, N. 2000. *Vocabolario della lingua italiana*. Bologna: Zanichelli.

A Lexico-phonetic Comparison
of Olukumi and Lukumi:
A Procedure for Developing
a Multilingual Dictionary

Joy Oluchi Uguru, *Department of Linguistics, University of Nigeria, Nsukka, Nigeria (joy.uguru@unn.edu.ng)*

Abstract: Generally, most multilingual dictionaries do not give adequate lexical and phonetic information (like contrasts and distributions). This could delay language learning (particularly among second language learners). This study demonstrates a comparative display of lexico-phonetic features of Lukumi and Olukumi in a proposed multilingual dictionary. The study, based on cognitive semantics and variation theories, proves that this display reveals how the user can distinguish the lexical and phonetic details within and across the languages. Downloaded Lukumi wordlists (132 words) were used to elicit information on Olukumi equivalents through an oral interview conducted in Ukwunzu, a major Olukumi speaking community in Delta state, Nigeria. However, 74 words were purposefully selected for comparative analysis while 23 words were used to demonstrate dictionary compilation. Through comparative analysis, free variants, synonymous and polysemous words were discovered and displayed in the dictionary. The study concludes that adequate lexical and phonetic comparison (and analysis) of words is vital in compiling a multilingual dictionary and will facilitate dictionary usage and language learning.

Keywords: LEXICO-PHONETIC, OLUKUMI, LUKUMI, MULTILINGUAL DICTIONARY, COGNITIVE SEMANTICS, VARIANTS, FREE VARIATION

Résumé: Une comparaison lexico-phonétique d'Olukumi et de Lukumi: une procédure pour développer un dictionnaire multilingue. En général, la plupart des dictionnaires multilingues ne donnent pas d'informations lexicales et phonétiques adéquates (comme les contrastes et les distributions). Cela pourrait retarder l'apprentissage des langues (en particulier chez les apprenants de langue seconde). Cette étude démontre un affichage comparatif des dispositifs phonétiques lexico de Lukumi et d'Olukumi dans un dictionnaire multilingue proposé. L'étude, basée sur la sémantique cognitive et les théories de la variation, prouve que cet affichage révèle comment l'utilisateur peut distinguer les détails lexicales et phonétiques dans et entre les langues. Les listes de mots Lukumi téléchargées (132 mots) ont été utilisées pour obtenir des informations sur les équivalents Olukumi grâce à une interview orale menée à Ukwunzu, une importante communauté parlant olukumi dans l'État du Delta, au Nigeria. Cependant, 74 mots ont été délibérément sélectionnés pour l'analyse comparative tandis que 23 mots ont été utilisés pour démontrer la compilation du dictionnaire. Grâce à l'analyse comparative, des variantes libres, des mots synonymes et polysémiques ont été découverts et affichés dans le dictionnaire. L'étude conclut qu'une comparaison (et une analyse) lexicales et phonétiques adéquates des mots est

essentielle à la compilation d'un dictionnaire multilingue et facilitera l'utilisation des dictionnaires et l'apprentissage des langues.

Mots-clés: LEXICO-PHONETIQUE, OLUKUMI, LUKUMI, DICTIONNAIRE MULTILINGUE, SÉMANTIQUE COGNITIVE, VARIANTES, VARIATION LIBRE

1.   Introduction

According to Rundell (2012) all linguistic procedures play important and key roles in dictionary compilation. This is so because all aspects of language are interconnected and these aspects, manifested through linguistic procedures, are displayed in the dictionary. Hence it is necessary to adopt the right linguistic procedures in order to have a good and reliable dictionary. In this paper, the procedures of phonetic transcription, procedures involving the determination of phonemic variants, procedures of parts of speech classification and procedures of meaning analysis through cognitive semantics are some of the procedures undertaken in the sample compilation of Lukumi and Olukumi (with English gloss) multilingual dictionary.

Schierholz (2015) shows that different methods and phases are involved in dictionary compilation. He cites Wiegand (1998) as outlining the following phases: the preparation phase, the phase of acquiring the material and the data, the phase of treating the material and the data; the evaluation phase and the phase of preparing the material for printing.

This study defines lexico-phonetic comparison as one of these phases that are necessary for compiling a good multilingual dictionary; it could be categorized under the phases of treating and evaluating the data. Lexico-phonetic comparison is important because the proposed dictionary is a new project; hence adequate information about the languages is necessary. According to Schierholz, the lexicographer should determine the type of project being undertaken (old or new). This will go a long way to help him or her know what steps to take.

Thus this study stands at a good pedestal to produce a reliable dictionary because the necessary linguistic procedures and lexicographic methods have been adopted. For such linguistic systems (as Lukumi and Olukumi) that are largely unwritten and without standard forms, good procedures are necessary to avoid producing a dictionary that may not effectively capture their lexicon. The display of pronunciation and phonetic variables in dictionaries on African languages, particularly, is rare (Uguru and Okeke 2020; Stark 1999). This study is geared towards performing this rare and difficult task since; presently, the use of Lukumi and Olukumi is mainly oral.

In Delta state of Nigeria, a number of linguistic systems with unique features abound. These include Ika which manifests intonation and tone (Uguru 2015) unlike other Igbo dialects. Olukumi is another unique system, being a Yoruboid language spoken in an environment where Igbo is predominantly spoken (in Oshimili Local Government Area). It has high similarities with Lukumi, also a

Yoruboid language spoken in Cuba (Uguru and Okeke 2020). Both varieties are in the New Benue Congo subgroup of the Niger Congo family. Lukumi has the code, ISO 639-3 luq while that of Olukumi (spelt Ulukwumi by Ethnologue) is ISO 639-3 ulb. Both are spoken by Yoruba descendants. These varieties resulted from slavery and migration respectively. Scholars have shown that Olukumi is highly related to Yoruba (Arokoyo 2012; Okolo-Obi 2014). Also, Lukumi is highly related to Yoruba (Ayoh'Omidire 2017).

Lukumi speakers are descendants of Yoruba slaves taken to Cuba. Olukumi is spoken by descendants of Yorubas who migrated from Western Nigeria to Eastern Nigeria; it is spoken in Ugbodu, Ukwunzu, Ubulubu, Idumu-ogo and Inyogo. See fig. 1 below.

Fig. 1:   Map of Olukumi speaking areas
Source: https://www.google.com.ng/search?q=map+of+olukumi+
speaking+areas+in+delta+state&tbm=isch&source=iu&ictx=1&fir=
hyLUQdP2J-

According to Mason (1997) in Cuba, Lukumi (also spelt as Lucumi, Ulcumi or Ulcami) refers to Africans of Yoruba descent as well as their language. He further reveals that in the United States, Lukumi is synonymous with *Orisha* worship because it is basically used for traditional Yoruba religion.

Both Olukumi and Lukumi are largely unwritten and not studied in schools. Therefore, the compilation of a dictionary is a good way of enhancing their development and usage for both oral and literary purposes. In this study therefore, we show how their lexical items can be compiled, displaying their lexical and phonetic features comparatively to enable language users and learners to easily capture their similarities and dissimilarities. This research therefore, will aid their documentation and preservation.

According to Mason (1997) Yoruba descendants are called by different names in various countries: in Brazil, they are known as the Nago or the Jeje while in Haiti they are known as the Nago. In Trinidad, they are called Sango/ Shango and in Cuba, they are known as the Lukumi. Hence Lukumi designates both the language and the Yoruba descendants.

Mason (1997) further shows that Lukumi was preserved due to Africans' resistance to whites' cultural oppression. It is closely tied to the Yoruba traditional religion. In fact, Mason (op cit.) shows that in the United States, Lukumi does not refer to the descendants of Africans from Nigeria or Cuba but rather to people (irrespective of ethnicity) who practise the Yoruba traditional religion.

According to Brandon (1993) Lukumi is a sacred ritual language. Santeria worshippers are forced to use Lukumi; many worshippers could not understand the language. Olmsted (1953) therefore shows that Lukumi is a highly conservative language; worshippers believe that they can go on with worship whether they understand the language or not. According to Brandon (1993) it is acquired in adolescence and adulthood, some people learning from copyings in notebooks; thus they have limited knowledge. Furthermore, Brandon reveals that the pronunciation and spelling of Lukumi are not uniform; the variety is known by several names like Lukumi, Ulcami, Lucumi and so on. This can be confusing. Hence, studying and documenting the language in a multilingual dictionary will aid in having uniform pronunciation and spelling for its name.

Ayoh'Omidire (2003) published a book, *Àkọ̀gbádùn: ABC da Língua, Cultura e Civilização Iorubanas,* for students in Brazil. It contains information about Orisha and Yoruba culture, poetries, songs and comparison of Brazilian Orishá traditions and Yorùbá customs. Arokoyo and Mabodu (2017) compiled Olukumi– English bilingual dictionary. However, it has some flaws as some of the lexical items were not well documented. For instance, *aṣọ* (cloth) was not documented yet *aṣọ abe* (under cloth), *aṣọ ẹrẹfù̀ná* (curtain) and *aṣọ oyin* (bee wax) were included. Also, though pronunciation was indicated, the IPA symbols were not strictly followed; this could be confusing to readers. Furthermore, phonetic features like allophonic variants were not shown. Most importantly, the information supplied is on one language; the English equivalents are just the gloss of the headwords. Also, Anderson, Arokoyo and Harrison (2012) compiled a talking dictionary. This also had lapses as many words were omitted in addition to the fact that the variants and other phonetic information necessary for effective language learning were not included.

There are concerns of imminent death of Olukumi due to the influence of Igbo language spoken in neighbouring communities (Onwueme 2015). Lukumi retains Yoruba features because it is solely used for religious purposes (Ayoh'Omidire 2017). Albeit, its survival is also threatened since it is not used for everyday communication; hence the necessity of compiling a dictionary on the two varieties. The lexico-phonetic comparison carried out in this study includes the procedures of phonetic and phonological analyses, classification of parts

of speech; and analysis of the meanings of lemmas. These will be aligned comparatively between the varieties and compiled for easy identification.

Lexico-phonetic information in dictionaries

While it is more straightforward, and perhaps, easier to reflect lexical information (particularly meaning and grammar) in dictionaries, it appears somewhat more tasking to indicate phonetic information of headwords in dictionaries (Sobkowiak 2000). This may be why most dictionaries, particularly those on African languages do not contain phonetic information (Uguru and Okeke 2020; Mbah et al. 2013). This is more so because the common goal in dictionary compilation is mainly meaning (Jain 2003).

The tendency for lexicographers to focus on meaning may be the reason why pronouncing dictionaries are necessary. In pronouncing dictionaries, the pronunciation and phonetic details of lemmas are displayed for language users. Such dictionaries display the pronunciation and variants to which language users can make reference. However, though these may be beneficial, they may not be easily available. Furthermore, they are not convenient to use; hardly could a language user obtain a pronouncing dictionary in addition to a learner's or general dictionary. Hence it is wiser to include phonetic information in the widely used types of dictionaries so that greater number of people will be conversant with the pronunciation of the language. Stark (1999) in his study of about fifty research works on dictionary usage, laments that only a few was centred on learners' pronunciation. Rather, the studies showed that learners neglect the phonetic aspect of language but are prone to looking up information on meaning, spelling and grammar. He summarizes this gap with the following excerpt from Sobkowiak (2000: 244):

> The place of phonetics in dictionaries generally, and in learners' dictionaries in particular, its role in the composition of the macro- as well as the microstructure of the dictionary, the wonder and challenge of multimedia in machine-readable dictionaries, the psycholinguistic issues of pronunciation look-up, and many others are all waiting to be researched.

Based on the foregoing, the present study recommends that the proposed Lukumi/Olukumi dictionary, which is a learner's dictionary, include phonetic information in addition to lexical/semantic information. This will enable foreign learners, native learners and other users have detailed information about the spoken form of the language. Hence, theories that can adequately account for, analyse and reflect both lexical and phonetic information about the varieties under study are necessary.

Booij (2003) laments that traditional dictionaries tend to emphasize written language. He shows that this ought to be corrected since information about the phonetic features of words is part of lexical information; hence it should not be ignored in the compilation of a dictionary. In giving the phonetic information,

the variants of phonemes (if any) are also displayed to enable the users of the dictionary to be conversant with all the available usage featuring in the linguistic system. For instance, a Yoruba dictionary with phonetic information should give the user information about the status of [ã] in Yoruba. It should show that it is not phonemically contrastive but rather in free variation relationship with [ɔ̃]. Jain (2003) has argued that each variation should be entered separately in the dictionary. We, however, argue against this because it will not only be confusing to the dictionary user, but will also make the document to be too voluminous. Rather, we propose that each variant should be attached to its headword. This way, the dictionary user knows alternating pronunciations to a given lemma. It has been shown that the pronunciation of a headword, given in International Phonetic Association symbols, should be clearly indicated in a dictionary (Jain 2003). The pronunciation of the symbols can also be simplified in the preliminary pages of the dictionary.

Sobkowiak (2000) reveals that there should be more research on the phonetic structure and choice of keywords so that the dictionary user can be well guided when looking up phonetic information. According to him, this task of phonetic look-up is difficult; hence the lexicographer needs to simplify it by making descriptions with the right choice of words and key phonetic structures. Also, words and phonetic transcriptions should be listed in such a way that they are easy to pronounce.

In giving lexical information in dictionaries, Booij (2003) opines that the emphasis should not be on giving all the possible meanings of a lemma; rather focus should be on showing its function and usage in the language. The part of speech of the lexical entry is also very vital information to be included in the dictionary. The meaning of the word, preferably given in one word, is the central information given about the lexical entry. It is however important to consider the type of dictionary and its users in giving information about headwords.

Theories for analyses

Linguistic and lexicographic theories enable the lexicographer to adequately analyse and synthesise linguistic data for dictionary compilation (Swanepoel 1994). The theories of variation and that of cognitive semantics form the base of this study. Cognitive semantics shows that language is acquired through cognition; that is, it is based on the conception of its speakers (how they conceive the world). Hence, each language will be made up of the concepts and objects around its speaker. Language is therefore, culture-bound. The concepts and objects first exist in the mind as thought patterns before being named. Hence it is only natural that people may not be able to have a lexical item for a concept or object that is not in their immediate environment, particularly if they do not have access to them. Thus though concepts are not tied to particular languages, they are influenced by environments and culture; this is why there are varied

categorizations of lexical items for concepts. Hence the data used for this study are concepts that Lukumi speakers are familiar with.

The variation theory enables us to examine, explain and link the variations between the linguistic varieties. Variation means saying the same thing in different ways (Meechan and Rees-Miller 2001). It means representing a concept or object with different words. It also includes the use of varying symbols in the lexical item to represent the same object or concept (Jarrar 2018). The latter definition is the main focus of this paper. Phonology appears to be a prominent domain in which variation features (Guy 2007).

Indicating linguistic variation in a dictionary will not only avail users with alternative usages, but also aid in informing them about the origins and etymology of the variable/variant. All languages have variations; these may originate from dialects, social groups, professions and so on. Hence linguistic variation is a natural phenomenon that should not be neglected in pursuit of a standard. In the case of Lukumi and Olukumi which are still undergoing development, without any standard, it is important to include their linguistic variables (particularly phonetic variables) in their documentation. This will aid in any future development of a standard form.

According to Lanwermeyer et al. (2016) dialect variation influences phonological and lexical-semantic word processing in sentences. In their lexico-phonological comparison of lexical items in two dialect areas (Central Bavarian and Bavarian-Alemannic transition zone) they discover that /oa /oa-oa-/ and /ʊ - ʊ -/ are used variously in the two dialects. For instance, the word for straw is /ʃtroa –oa -/ in BA, but in CB it is pronounced as /ʃtroʊ-ʊ-/. This kind of variation, if not explained to language learners, could lead to difficulties in form–meaning associations in the dialect areas. Hence a dictionary such as the one this study projects, is very necessary. When these sound variations do not yield meaning differences, then the varying sounds are allophones and this must be pointed out in the dictionary. Information to be given includes whether occurrence is conditioned as well as the environments in which they occur. If they are not conditioned, then they are free variants (in free variation). Free variation is a situation where two or more sounds or forms occur in the same environment without a change in meaning. Alternating variants occur in regular patterns (Guy 2007).

Hence, it is significant that Olukumi and Lukumi, which are spoken by people in different continents, have a lot of lexical and phonetic similarities. Going by the cognitive semantic interpretation of their word meanings, lexical and phonetic similarities portray the fact that they have the same origin. There can be no other plausible explanation for this high degree of similarities.

Hence it is important to indicate the phonetic features of words in a dictionary since it will not only make for ease of usage, but will also reveal the relationship (that is, similarities or otherwise) between the concerned languages. Dellert et al. (2020) assert that most grammar books contain general information about the phonology of languages, and that languages, particu-

larly less documented ones, rarely have phonetic transcriptions for individual words. They show that this causes people to depend on the written forms which could be problematic if the orthography does not fully represent pronunciation.

Adda-Decker and Lamel (2006) reveal that even in the use of speech recognizing systems, indicating phonetic features in multilingual dictionaries, particularly, helps to reduce poor performance since multilingual dictionaries contain non-native speech forms. However, most lexicographers de-emphasize phonetic features and that has resulted in its non-reflection in most dictionaries. Čermák (2010) in discussing steps in dictionary compilation, de-emphasizes pronunciation, showing that it is used only for distinction and for foreign words. This view is erroneous since anybody can benefit from the indication of the phonetic features of entries in any dictionary because both native and second language speakers (and also learners) can have access to the dictionary.

2.   Methods

Lukumi word lists were downloaded from some websites (cf. references) because Lukumi is not spoken in the environment of research. Seventy four words were purposefully selected for comparative analysis and twenty three words were used to demonstrate dictionary compilation. An Olukumi native speaker supplied the Olukumi equivalents of the downloaded words during an oral interview. The equivalents in Olukumi and Lukumi were compared with words bearing similar concepts. Cognitive theory was used to analyse lexical meaning and variation theory was used to determine the phonetic variables in the two varieties. The effect of these phonetic variables on the meanings of lemmas was examined in the varieties.

Additionally, similarities and differences in the occurrence and distribution of Lukumi and Olukumi phonemes were determined. Based on this, a sample compilation of the proposed dictionary was done.

3.   Lexico-phonetic comparison of Lukumi and Olukumi

In this section, the words are analysed in terms of their meanings and parts of speech and the phonemes are analysed in terms of their similarities, variation and distribution.

3.1   Phonetic comparison of the speech sounds of Lukumi and Olukumi

In this section, phonemes and lexical items that made up the sample wordlist are displayed in the following tables.

Table 1:   Speech sounds of Lukumi, Olukumi and the impacting languages

Language	Plosives	Fricatives	Affricates	Nasals	Flap	Trill	Lateral	Approxi-mant	Vowels
Olukumi	b t d k g kp gb gʷ	f s z ʃ ɣ ɦ	dʒ	m n ɲ ŋ ŋʷ		r	l	j w	i u ũ ɪ ĩ ʊ e o ɛ ɛ̃ ɔ õ a ã
Igbo	p b t d k g kp gb kʷ gʷ	s z f ʃ (ʒ) ɣ ɦ	tʃ dʒ	m n ɲ ŋ ŋʷ			l	ɹ j w	i ɪ u ʊ o ɔ e ɛ a
Lukumi/ Anago	b gb kp d t g k	f s	tʃ	m n ɲ ŋ		r	l	j	i e ɛ a o u
Spanish	p b t d k g	f θ s ʃ j x	tʃ	m n ɲ	ɾ	r	l ʎ		i u e o a
Yoruba	b t d ɟ k g kp gb	f s ʃ h	dʒ	m n ŋ		r	l	j w	i ĩ u ũ e o ɛ ɛ̃ ɔ õ a

Note: Phonemes are placed without regard to voiced/voiceless positions

Phonetic comparison of Lukumi and Olukumi

Phonetic similarity is concerned with articulatory, acoustic, and perceptual similarities between vowels and consonants (Schepens et al. 2013).

In terms of the phonetic details, it can be observed that some Yoruba phonemes which do not exist in Lukumi exist in Olukumi. Hence Lukumi speakers replace the phonemes with those nearest to them in articulation. For instance /gb/ is replaced with /b/ in Lukumi since the former does not feature in the variety.

3.2   Prosody

The analysis of the prosodic features in the varieties is shown below.

3.2.1 Tone

Adeshokan (2018) reveals that just like in Spanish, accents feature in Lukumi words. Tone, a feature of Yoruba, the parent language, does not feature in Lukumi but it exists in Olukumi.

3.2.2 Stress and syllabic structure

Lukumi words are not tone marked but rather accents (typical of Spanish) are used to indicate stressed syllables and they usually occur in word final position (Ramos

2012; Concordia 2012). As can be observed from the data, Lukumi and Olukumi have CV syllable structure. Hence they do not have closed syllables. Spanish has a closed syllable structure but that did not influence Lukumi.

Also there are some linguistic processes that occur in both Lukumi and Olukumi varieties. For instance, syllabic repetition as a way of expressing colour is evident in the expression of *dudu* (Lukumi for dark) and *okwukwu* (Olukumi for the same colour). Similarly, *funfun* (white) for both varieties has syllabic repetition. Furthermore, syllable elision is observed in Olukumi. Observe the examples below.

Lukumi	Olukumi
Baba	ba
Babalawo	awo
Yeye	ye

3.3   Lexical comparison of Lukumi and Olukumi

Vowel nasality appears to be a common feature in both varieties as seen in *funfun, nwun, eyin* and so on. Thus the varieties share some phonetic features in addition to lexical and semantic similarities.

These can be clearly shown in lexical items in Lukumi and Olukumi which appear below.

Table 2:   Lexical items in Lukumi and Olukumi

Lukumi	Olukumi	English
1. Abó /abo/	Àgbò /agbo/	Ram
2. Adié /adie/	Eduwe /eduwe/	Hen
3. Agogó /agogo/	Agogo /ágógó/	Bell, Hour
4. Agoya /agoja/	Yà /ja/	Enter
5. Akukó /akuko/	Ákị́kọ́ /akɪkɔ/	Rooster
6. Ala /ala/	ɛ̀lá /ɛla/	Dream
7. Aná /ana/	ɔ̀nà /ɔna/	Road
8. Ara /ara/	Àkpàrà /akpara/	Thunder
9. Ashelú /aʃelu/	Efedudu /efedudu/	Police
10. Ashere /aʃere/	Ero /ero/	Maraca, Rattle
11. Asho /aʃɔ/	áṣọ́ /aʃɔ/	Cloth
12. Ayá /aja/	àzá /aza/	Dog
13. Babá /baba/	bá /bá/	Father
14. Babalawo /babalawo/	àwó /awo/	Father of the Secrets/herbalist
15. Busi /busi/	Gọzi /gɔzi/	Bless
16. Ejá /eʤa/	ézá /eza/	Fish

17. Eje /ɛʤɛ/	èzè /ɛzɛ/	Blood
18. Ejo /eʤo/	é'zó /ɛzɔ/	Court case
19. Ekpó /ekpo/	Ékpó /ekpo/	Palm oil
20. Ekú /eku/	Egugu /eɡuɡu/	Masquerade regalia
21. Eñe /eɲe/	Ényí /eɲi/	Tooth
22. Ení /eni/	Éní /eni/	Mat
23. Funfún /fũfũ/	Fúnfún /fũfũ/	White
24. Gidigidi /ɡidiɡidi/	O lala /o lala/	Very Much
25. Ibú /ibu/	Omi /omi/	Stream, River
26. Igba /iɡba/	Ụgban /ʊgbã/	Calabash
27. Ikú /iku/	Úkú /uku/	Death
28. Ilé /ile/	Úlé /ule/	House
29. Ilekún /ilekun/	èkùn /ɛkʊn/	Door
30. Iñá /iɲa/	Ụ́ná /ʊna/	Fire
31. Iré /ire/	Íré /ire/	Blessings/gift
32. Ishu /iʃu/	Ú'sún /usun/	African yam
33. Iworo /iworo/	Awo /awo/	Priest
34. Juba /ʤuba/	Ushi /ushi/	Salute
35. Kunlé /kunle/	Kụale /kụale/	Kneel
36. Leti /leti/	étí /eti/	Ear
37. Lo /lo/	Are /are/	Go
38. loyú /loju/	ózú /ozu/	Eye
39. Maferefún /maferefũ/	Ụchira wu /ụchi ra wu/	Praise be to
40. meyi /meji/	mêzìn /mezin/	Two
41. Mi /mi/	Témí /temi/	My
42. Moducué /moduke/	Ese /ese /	Thank you
43. Mojuba /moʤuba/	Moshien /moshien/	I salute you
44. Nlo /nlo/	Are /are/	Is going
45. Obá /oba/	Ọlọza /ɔlɔza	King
46. Obí /obi/	Óbì /obi/	Kola nut
47. Odo /odo/	Omi /omi/	River
48. Ofo /ofo/	Ono /ono/	Loss
49. Oguede /oɡuede/	ògèdè /ɔgɛdɛ/	Plantain
50. Oke /oke/	èdòkè /edoke/	Mountain
51. Oko /oko/	Ọ́kọ́ /ɔkɔ/	Man, Husband
52. Okún /okũ/	Omi lila /omi lila/	Sea, Ocean
53. Okuta /okuta/	Okuta /okuta/	Stone
54. Olóun /oloun/	Enune /enune/	Owner, on who has...
55. Olúwo /oluwo/	Enune wa /enune wa/	Chief priest; Lord of Awos (Babalawo who became an orisha priest)
56. Omí /omi/	Ómí /omi/	Water
57. Omí Dudu /omi dudu/	Ómí ụ́ná /omi ʊna/	Coffee

58. Omo /ɔmɔ/	Ọ́má /ɔma/	Child
59. Ona /ona/	ọ̀nà /ɔna/	road
60. Oni /oni/	Enune /enune/	Owner of ...
61. Opolopo /okpolokpo/	Lala /lala/	Plenty
62. Orí /ori/	Ọ́rí /ori/	Head
63. Orí /ori/	Okuma /okuma/	Cocoa butter
64. Orún /orũ/	Orúnrún /orũrũ/	Sun
65. Otí /oti/	Omi ụna /omi una/	Rum
66. Owú /owu/	Òwú /owu/	Cotton
67. Oyín /ojĩ/	ómí ónyí /omi ọni/	Honey
68. Pupua /pupua/	Kpíkpán /kpikpan/	Red
69. Shishé /ʃiʃe/	úsé /use/	Work
70. Surefun /surefũ/	Gozi /gɔzi/	Bless (bless him/her)
71. Tobí /tobi/	Óbi /obi/	Who gave birth
72. Tutu /tutu/	títọ́n /tɪtɔn/	Cool, fresh
73. Wa /wa/	Wa /wa/	Come
74. Yeye, /jeje/ Iyá	yé /je/	Mama/mother

From the data above, some distribution of phonemes can be seen. These are shown in table 3.

Table 3:   Distribution of phonemes in Lukumi and Olukumi

Phoneme(s)	Lukumi	Olukumi	English
b, gb	Abo /abɔ/	Agbo /agbo/	ram
ʤ, z	eja /eʤa/ eje /eʤe/ ejo /eʤo/	eza /eza/ ẹzẹ /ɛza/ é⁺zọ́ /ɛzɔ/	fish blood court case
u, ɪ	akuko /akuko/	akịkọ /akɪkɔ/	rooster
i, u; ʊ	ile /ile/	ule /ule/	house
j, z	Loyu /loju/	Ozu /ozu/	eye
ʃ, s	shishé /ʃiʃe/ ishi /iʃi/	use /use/ usu /usu/	work African yam

3.4   Discussion

From Tables 2 and 3, it can be observed that many of the concepts in the two varieties have the same form units. Also, some allophones exist in the varieties.

That is, some words have varying phonemes without such variations yielding changes in meaning. The allophones can be proposed to be used as free variants by language users/learners since the lemmas do not have any change of meaning when the allophones are interchanged; more so, their occurrence does not appear to be conditioned by any environment/phoneme. This is similar to what operates in the English language where either British or American spellings can be used by writers. Just as American spelling and pronunciation are indicated in dictionaries of English language, and users are free to use any one, in the proposed dictionary, Lukumi and Olukumi allophonic variations are indicated for users. Consequently, dictionary users are free to adopt anyone they feel like using since the lexical meanings are not changed by the use of varied allophones. This is what this study is all about: to display these allophonic variations and show their impact on lexical meanings.

Indicating the free variation existing between the two varieties will go a long way to help speakers and dictionary users alike. This is so because, since the variations are not complementary, grammar books alone cannot adequately be used to explain their usage to users of the languages. The variations must necessarily be pointed out in the dictionary and this is what the projected dictionary will do. This will go a long way to help learners, particularly those who use Lukumi mainly for religious purposes.

The meaning analysis of the words used for the study portrays that there are some synonymous and polysemous words in the varieties. It takes lexical analysis of meaning, using cognitive semantics, to arrive at this. These are pointed out for learners as shown in our sample compilation below. The Olukumi word for water *omi* for instance, is used for *water, stream/river;* and indicates it in its equivalent for ocean/sea 'omi lila' *ocean*. On the contrary, furthermore, the word 'omi' is used to show a sense of liquid as seen in *omi una* (rum; Olukumi data 65) (hot drink – literal translation). The same expression is also used synonymously with the equivalent for coffee (see data no. 57). *Omi dudu* (Lukumi data 57) and *efe dudu* (Olukumi data 9) reflect that both varieties make use of description in naming some concepts. Lukumi has the following words for these concepts: *omi,* (water), *ibu* (stream), *okun* (ocean). Similarly, Lukumi has a number of synonyms as shown in Table 2. While Olukumi has only one word, *gozi,* for *bless,* in Lukumi, both *busi* and *surefun* mean *bless.* All these could be confusing for a language learner hence it should be the duty of the lexicographer to reduce ambiguity by accurate indication of these features in the dictionary. These can be seen in the sample compilation below.

4.   Sample compilation of lemmas for the multilingual dictionary

In a previous paper, a certain form of compilation was adopted because three languages were involved; also, free variants and detailed lexical information as well as cross references were not included. An excerpt from that paper is shown below (Uguru and Okeke 2020).

Word	Pronunciation	Language	English translation
abo	/abo/	n. LUK	Ram
àgbò	/agbo/	n. OLU	
àgbò	/agbo/	n. YOR	

Hence, still maintaining space economy as was done in the previous paper, a new compilation method is adopted here to create room for more information that this study sets out to display in the dictionary.

In the proposed dictionary, it will be clearly pointed out at the preliminary pages, that since most head words have the same meanings in both varieties, indication of variety will only be made in cases of sound and meaning differences. That is, where there is a difference in form units. The preliminary pages will also contain the phonemes of the varieties as well as the sounds involved in free variation. Below, demonstration is made, showing sample compilation of the dictionary. The compilation shows how lemmas with the same form units can be entered, how those with varying phonemes can be entered, and how those that differ in form units can be represented in the dictionary. In addition, lexical and meaning information is given, pointing out synonymous and polysemous headwords.

4.1   The Sample Dictionary Compilation

4.1.1   Preliminary page information

Outline of phonemic distributions: In most cases, sound distribution in both varieties is as follows: in most environments where Lukumi would use the following sounds — /b, e, ʤ, i, j, ʃ/ Olukumi would use the following: /gb, ɛ, z, u, z and s/. Hence the following sounds could be used interchangeably, in the varieties, for some headwords which have the same meaning but vary in one or two phonemes: b/gb, e/ɛ, ʤ/z, i/u, j/z and ʃ/s.

1.   **Abo** /abo/ (ram) N; **àgbò** /agbo/ (Oluk)
     [b] / [gb] (free variants)

2.   **Agogó** /agogo/ (bell) N.

3.   **Akuko** /akuko/ (rooster) N. **aki̧ko̧** /akɪkɔ/ (Oluk)
     u / ɪ (free variants)

4.   **Babalawo** (herbalist; keeper of secrets) N; **Awo** /awo/ (Oluk): OLUWO; IWORO

5.   **Busi** /busi/ (bless) V; **gọzi** /gɔzi/ (Oluk): SUREFUN

6.   **Eja** /eʤa/ (fish) N; **eza** /eza/ (Oluk)
[ʤ] / [z] free variants

7.   **Eje** /eʤe/ (blood) N; **ẹzẹ** /ɛzɛ/ (Oluk)
[e] / [ɛ]; [ʤ] / [z] free variants

8.   **Ejo** /eʤo/ (court case) N; **ẹ⁺zọ** /ɛzɔ/ (Oluk)
[e] / [ɛ]; [ʤ] / [z] free variants

9.   **Ibú** /ibu/ (stream, river) N; **omi** /omi/ (Oluk) — COMPARE OKUN, ODO, OMI

10.   **Ile** /ile/ (house) N; **ule** /ule/ (Oluk)
[i] / [u] free variants

11.   **Ishi** /iʃi/ (African yam) N; **usu** /usu/ (Oluk)
[i] / [u]; [ʃ] / [s] free variants

12.   **Iworo** (chief priest) N; **awo** /awo/ (Oluk): BABALAWO, OLUWO

13.   **Iyá** /ija/ (fight) N; **ùzà** /uza/ (Oluk)
[i] / [u]; [j] / [z]; free variants

14.   **Loyu** /loju/ (eye) N; **Ozu** /ozu/ (Oluk)
[j] / [z] free variants

15.   **Odo** /odo/ (river) N; **omi** /omi/ (Oluk): COMPARE IBU, ODO, OMI

16.   **Okun** /okũ/ (ocean) **omi lila** /omi lila/ (Oluk): COMPARE IBU, ODO, OMI

17.   **Oluwo** /oluwo/ (chief priest, lord of Awos) N; **Enune wa** (Oluk): BABALAWO, IWORO

18.   **Omi** /omi/ (water) N.

19.   **Owó** /owo/ (money) N. **éghó** /eɣo/ (Oluk)

20.   **Pupua** /pupua/ (red) Adj; **Kpíkpán** /kpikpan/ (Oluk)
[p] / [kp] free variants

21.   **Shishé** /ʃiʃe/ (work) N; **use** /use/ (Oluk)
[i] / [u]; [ʃ] / [s] free variants

22.   **Surefun** /surefũ/ (bless him) V; **gọzi** /gɔzi/ (Oluk): BUSI

23.   **Tutu** /tutu/ (cool, fresh) Adj.; **tįtọ́n** /tɪtɔn/ (Oluk)

Key of abbreviations used in the sample:
Adj. — adjective
N. — noun

Oluk. — Olukumi
Key of abbreviations:
Adj — adjective
N — noun
Oluk. — Olukumi
V — verb

Many phonemes, words and concepts in Lukumi and Olukumi are similar despite the distance separating the locations where they are spoken. Above, we have shown these lexical and phonetic similarities in a sample Lukumi–Olukumi multilingual dictionary (with English gloss). For the entries where free variants have been indicated for instance, either of the pronunciation is acceptable as seen in *economics* in English which can be pronounced /eknomiks/ or /ɪknomɪks/. It is also applicable to Igbo language where the word for *ground* can be pronounced as /ala/, /alɪ/, /ana/ or /anɪ/. Hence the pronunciation for the word for *ram* can either be pronounced as /abo/ or /agbo/. As the features of these varieties are maintained this way, they will not die; this will particularly benefit the Nigerian variety, Olukumi, whose existence is said to be threatened (Onwueme 2015).

4.2   Discussion

There are lots of similarities in the phonemes and lexical items of Lukumi and Olukumi. Interestingly, lexical similarity appears to align with phonetic/phonemic similarity in these varieties. They have mainly phonetic spelling; that is, most of their phonemes bear the same symbols as the letters of their alphabet. This is because according to Coulmas (1996) alphabets for African languages were influenced by the work of phoneticians at the International Institute of African Languages and Cultures in London. They established the Practical Orthography of African Languages; it was influenced by the International Phonetic Association, thus being based on the principle of one letter corresponding to one sound.

4.2.1   Implications of the phonetic and lexical similarities for a multilingual dictionary

Showing the phonetic and lexical similarities of entries, as done in our sample compilation in this paper, equips the dictionary user (especially language learners) adequately to undertake the use of the dictionary with ease, being able to distinguish the sounds and entries that are peculiar to a variety and the ones they share. This reduces errors in language use.

Indicating the phonetic components of words in a dictionary makes the dictionary effective for comprehension (reading and listening) and production (writing and speaking) Mdee (1997).

4.3   The significance of the multilingual dictionary in Lukumi and Olukumi conservation

In line with Kroskrity (2015) this work has documented the lexical items of two varieties, Olukumi and Lukumi and compiled them, providing their English gloss. The study reveals their lexical and phonetic features and displays these in the sample dictionary compiled in section 4.1. This will help to preserve Olukumi (spoken in Delta state, Nigeria) which is largely endangered as well as Lukumi (spoken in Cuba) which is not used for everyday activities but rather solely for religious purposes. This step will preserve these varieties, spoken in diaspora, from going into extinction. Dictionaries are of obvious importance to endangered language communities, being learning resources to speakers, including those who are acquiring their heritage languages as second languages (Haviland 2006: 129).

The use of Lukumi solely for religious purposes cannot guarantee its maintenance. For instance the sole use of Latin as a religious language has not enhanced its survival or evolution into a modern language. Hence the documentation and compilation of these varieties (alongside English translations in a multilingual dictionary) as exemplified in this paper, will go a long way to ensure their regular and wider usage, thereby preserving them from imminent death.

5.   Conclusion

In this paper, we have been able to comparatively analyse the lexico-phonetic features of Lukumi and Olukumi. The analysis aided the display of lexico-phonetic features in a sample compilation of a multilingual dictionary on the varieties. Thus, lexical and phonetic information of dictionary entries were shown in the dictionary, while maintaining space economy. It was discovered that the varieties have many concepts that are represented with the same form units; hence their lexical similarities are much. In the same vein, a lot of their phonemes are similar though there are a few that are peculiar to either varieties and these peculiar ones tend to occur in the same environment with their counterparts in the other variety. Due to this peculiarity, they share many words that are in free variation. This is already shown in the sample. Also, there are synonymous and polysemous words in the varieties, particularly Lukumi.

Hence, in the sample compilation, all these were taken into account. Information about the entries in free variation was included. Furthermore, the entries that are synonymous and polysemous respectively, were indicated through cross-referencing. With this depth of information given in the dictionary, users will find it easy to understand the varieties; language learning will be a lot facilitated.

The study confirms that dictionary compilation should not be haphazard. A good analysis of the language to be compiled is important to arm the lexicog-

rapher with detailed information to display in the dictionary. Based on the analysis of the two varieties under study, the entries of the proposed dictionary are categorised as follows: words of the varieties that have the same form units and the same meaning; those that have different form units for the same concept; those that differ in one or two phonemes (free variants) across the varieties and those concepts that are represented by more than one word as well as some words that denote more than one concept. Hence this knowledge enabled adequate explanation of the entries. Therefore, we have shown how the lexical items can be compiled in the dictionary in such a way that the dictionary user can easily identify features that distinguish the entries, those that the varieties share as well as those they can interchange (free variation). This has been made possible by the analysis of lexical and phonetic features of the words.

References

Dictionaries and word lists

Anderson, G.D.S., B. Arokoyo and K.D. Harrison. 2012. *Olùkùmi Talking Dictionary.* Living Tongues Institute for Endangered Languages. Available at: http://www.talkingdictionary. org/olukumi and http://talkingdictionary.swarthmore.edu/olukumi/.

Arokoyo, B.E. and O. Mabodu. 2017. *Olùkùmi Bilingual Dictionary*. 2017. Oregon: Living Tongues Institute for Endangered Languages.

Lucumi Dictionaries. Available at: https://www.orishaimage.com/blog/dictionaries. Downloaded 30 January 2018.

Lucumi Vocabulary. Available at: http://www.orishanet.org/vocab.html. Downloaded 30 January 2018.

Map of Olukumi speaking areas

Available at: https://www.google.com.ng/search?q=map+of+olukumi+speaking+areas+in+delta+ state&tbm=isch&source=iu&ictx=1&fir=hyLUQdP2J-. Downloaded 30 January 2018.

Oral interview

Lucumi Vocabulary. Available at: http://www.orishanet.org/vocab.html and https://www. orishaimage.com/blog/dictionaries. Downloaded 30 January 2018.

Ogwu, E. 2017. Word list collected from Ogwu, E. 2017. Headmaster of a primary school in Ukwunzu, (oral interview) researcher: 31st March 2017.

Other literature

Adda-Decker, M. and L. Lamel. 2006. *Multilingual Dictionaries.* Schultz, T. and K. Kirchhoff (Eds.). 2006. *Multilingual Speech Processing*: 123-168. Amsterdam et al.: MA: Academic Press.

Adeshokan, O. 2018. An In-depth Look into Lucumi and Yoruba in Comparison. *The Guardian,* 18 March 2018. Available at: https://guardian.ng/life/an-in-depth-look-into-lucumi-and-yoruba-in-comparison/.

Arokoyo, B.E. 2012. A Comparative Phonology of the Olùkùmi, Igala, Owe and Yoruba Languages. Paper presented at the Conference on *Towards Proto-Niger Congo: Comparison and Reconstruction*, Paris, 18–21 September 2012.

Ayoh'Omidire, F. 2003. *Àkọ̀gbádùn: ABC da língua, cultura e civilização iorubanas.* (ABC of the Yoruba language, Culture and Civilization.) Salvador: EDUFBA.

Ayoh'Omidire, F. 2017. *Yorùbá, Lukumí and Nagô: The Ilé-Ife Perspective*. Available at: http://www.orishaimage.com/blog/felixayohomidire. Accessed 12/08/18.

Booij, G. 2003. The Codification of Phonological, Morphological, and Syntactic Information. Van Sterkenburg, P. 2003. *A Practical Guide to Lexicography*: 251-259. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/tlrp.6.30boo. Available at: https://www.researchgate.net/publication/300848849 6.1.

Brandon, G. 1993. *Santeria from Africa to the New World: The Dead Sell Memories.* Bloomington: Indiana University Press.

Čermák, F. 2010. Notes on Compiling a Corpus-Based Dictionary. *Lexikos* 20: 559-579.

Concordia, M.J. 2012. *The Anagó Language of Cuba.* Unpublished M.A. Thesis. Miami: Florida International University.

Coulmas, F. 1996. *The Blackwell Encyclopedia of Writing Systems*. Oxford, UK/Cambridge, Mass.: Blackwell.

Dellert, J., T. Daneyko, A. Münch, A. Ladygina, A. Buch, N. Clarius, I. Grigorjew, M. Balabel, H.I. Boga, Z. Baysarova, R. Mühlenbernd, J. Wahle and G. Jäger. 2020. NorthEuraLex: A Wide-coverage Lexical Database of Northern Eurasia. *Language Resources and Evaluation* 54: 273-301.

Guy, G.R. 2007. Variation and Phonological Theory CUUK1061B-Bayley, R. and C. Lucas (Eds.). 2007. *Sociolinguistic Variation: Theories, Methods, and Analysis:* 5-23. Cambridge: Cambridge University Press. Available at: http://gregoryrguy.com/wp-content/uploads/GuyProofs-BayleyLucasvol.pdf.

Haviland, J.B. 2006. Documenting Lexical Knowledge. Jost, G., P.H. Nikolaus and M. Ulrike (Eds.). 2006. *Essentials of Language Documentation:* 129-162. Berlin: Mouton de Gruyter.

Jain, M. 2003. *Lexicography for Endangered Languages*. Available at: http://www.tezu.ernet.in/wmcfel/pdf/Cog/lexico/03.pdf.

Jarrar, M. 2018. *Introduction to Lexical Semantics.* Lecture Notes on Introduction to Lexical Semantics. Birzeit University, Palestine. Available at: http://www.jarrar.info/courses/Jarrar.LectureNotes.LexicalSemantics.pdf.

Kroskrity, P.V. 2015. Designing a Dictionary for an Endangered Language Community: Lexicographical Deliberations, Language Ideological Clarifications. *Language Documentation and Conservation* 9: 140-157.

Lanwermeyer, Manuela, Karen Henrich, Marie J. Rocholl, Hanni T. Schnell, Alexander Werth, Joachim Herrgen and Jürgen E. Schmidt. 2016. Dialect Variation Influences the Phonological and Lexical-Semantic Word Processing in Sentences. Electrophysiological Evidence from a Cross-Dialectal Comprehension Study. *Frontiers in Psychology* 7:739. https://doi.org/10.3389/fpsyg.2016.00739.

Mason, J. 1997. Ogun: Builder of the Lukumi's House. Barnes, S.T. (Ed.). 1997. *Africa's Ogun: Old World and New:* 353-368. Second, expanded edition. Bloomington/Indianapolis: Indiana University Press.

Mbah, B.M., E.E. Mbah, E.S. Ikeokwu, C.O. Okeke, I.M. Nweze, C.N. Ugwuona, C.M. Akaeze, J.O. Onu, E.A. Eze, G.A. Prezi and B.C. Odii. 2013. *Igbo Adi.* Nsukka: University of Nigeria Press.

Mdee, J.S. 1997. Language Learners' Use of a Bilingual Dictionary: A Comparative Study of Dictionary Use and Needs. *Lexikos* 7: 94-106.

Meechan, Marjory and Janie Rees-Miller. 2001. Language in Social Contexts. O'Grady, William, John Archibald, Mark Aronoff and Janie Rees-Miller (Eds.). 2001. *Contemporary Linguistics:* 485-524. Fourth edition. Bedford: St. Martin's.

Okolo-Obi, B. 2014. *Aspects of Olukumi Phonology.* A Project Report of the Department of Linguistics, Igbo and other Nigerian Languages, University of Nigeria. Nsukka: University of Nigeria.

Olmsted, D.L. 1953. Comparative Notes on Yoruba and Lucumí. *Journal of the Linguistic Society of America* 29 (2): 157-163.

Onwueme, I.C. 2015. *Questions Not Being Asked: Topical Philosophical Critiques in Prose, Proverbs, and Poems.* Bloomington: AuthorHouse.

Ramos, M.W. 2012. *Obí Agbón: Lukumí Divination with Coconut.* Miami: Eleda.Org. Available at: https://books.google.com.ng/books?isbn=1877845116. Downloaded 27 January 2018.

Rundell, Michael. 2012. It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012, Oslo:* 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Schepens, J., T. Dijkstra, F. Grootjen and W.J.B. van Heuven. 2013. Cross-language Distributions of High Frequency and Phonetically Similar Cognates. *PLoS One* 8(5): e63006.

Schierholz, Stefan J. 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.

Sobkowiak, W. 2000. Phonetic Keywords in Learner's Dictionaries. Heid, U. et al. (Eds.). 2000. *Proceedings of EURALEX 2000*: 237-246. Stuttgart: IMS, Universität Stuttgart.

Stark, M. 1999. *Encyclopedic Learners' Dictionaries.* Tübingen: Max Niemeyer.

Swanepoel, P. 1994. Problems, Theories and Methodologies in Current Lexicographic Semantic Research. Martin, W. et al. (Eds.). 1994. *Euralex 1994. Proceedings, Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 11-26. Amsterdam: Vrije Universiteit.

Uguru, J.O. 2015. Ika Igbo. *Journal of the International Phonetic Association* 45(2): 213-219.

Uguru, J.O. and C.O. Okeke. 2020. Reflecting Pronunciation in a Multilingual Dictionary: The Case of Lukumi, Olukumi and Yoruba Dictionary. *Lexikos* 30: 519-539.

Wiegand Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Volume 1.* Berlin/New York: Walter de Gruyter.

Sarah Ogilvie and Gabriella Safran (Eds.). *The Whole World in a Book: Dictionaries in the Nineteenth Century.* 2020, 358 pp. ISBN: 978-0190913199 (Hardback). Oxford: Oxford University Press. Price: £22.99.

Dictionaries are works of individual lexicographers or a group of lexicographers and also constrained by social and cultural factors of the time. Thus, they bear unmistakable features of the makers and the period. Historical studies of dictionaries can reveal the facts about the dictionaries and their compilers, and it is necessary to carry out such studies to promote further progress in academic lexicography (Hartmann 2001: 39).

The nineteenth century witnessed the dramatic development of industrialization and globalization. The industrial revolution and technological development opened up increasing opportunities for the mobility of people and sped up communication, which brought about the burgeoning of dictionaries worldwide. Dictionaries in the nineteenth century bore features of the time: they promoted national identity; they reflected different views on language; they portrayed authorship differences and lexicographic innovations. Previous historical studies of dictionaries mainly focused on different versions of a dictionary or dictionaries in one country. A more recent exception is *The Cambridge World History of Lexicography* edited by John Considine (2019). The present book can be seen as an addition, exploring dictionaries in the nineteenth century worldwide and giving answers to the following four questions: Who were the nineteenth-century lexicographers? How did the world within which they lived foster their projects? What did language itself mean to them? What goals did they try to accomplish in their dictionaries? (p. xv). The book is divided into 16 chapters together with an introduction and an index. The articles are organized on the principle that the dictionary dealt with is from the general to the specific, and from the more familiar to the less familiar (p. xix). The dictionaries include famous ones such as *Oxford English Dictionary* (OED), Noah Webster's *American Dictionary of the English Language*, Vladimir Dahl's *Explanatory Dictionary of the Living Great-Russian Language,* etc., but also less known Chinese, French, Frisian, Japanese, Persian, Scottish and Yiddish dictionaries, and dictionaries of sign language.

In Chapter 1, John Considine provides a general background of nineteenth-century lexicography by pointing out the challenges needing to be addressed by the lexicographers. Despite the flourishing of lexicography in the eighteenth century, some problems remained unsettled. The expanding and speeding up of communication also brought up new questions concerning the structure and contents of the dictionary entry, the scope of the dictionary wordlist, and the type of information which dictionaries should offer. Considine points out that the questions must be considered 'in the context of the whole ecosystem of reference publishing' (p. 13).

In Chapter 2, Brian Kim traces the development of foreign language dictionaries and dictionaries of the native language in the nineteenth century in

Russia and Japan with statistical evidence. The increasing tendency was the result of greater contact with Western Europe and the efforts to seek modernization while maintaining the national identity.

In Chapter 3, Michael Adams describes and evaluates Charles Richardson's *New Dictionary of the English Language* (1836–1837). As a mirror of the compiler's philosophy of language, NDEL is a valuable foil in the history of lexicography and is 'the pivot on which the scientific revolution of English lexicography turned' (p. 49).

In Chapter 4, Sarah Ogilvie explains why OED features prominently in modern lexicography. It was a collaborative work of both specialists and the general public. It applied historical principles systematically to the structure and content of the dictionary entry, adopted a descriptive rather than prescriptive approach, attempted a thorough coverage of the lexicon, and was received as a national project.

Chapter 5 deals with the *Deutsches Wörterbuch* (DWB), which is considered the starting point of scientific lexicography. The scientific feature was illustrated by the descriptive agenda, the description of word meaning, the collaborative network, data collection, etc. Volker Harm also shows that its scientific approach was still deeply rooted in Romanticism.

In Chapter 6, Anne Dykstra studies Joost Hiddes Halbertsma's unfinished dictionary of Frisian, the *Lexicon Frisicum* (1872). The dictionary reveals strong nationalistic tendencies of Romanticism. Dykstra analyzes it in terms of structure, source and material, metalanguage, meaning description, examples, etymology, and encyclopedic information, as well as the relationship between Frisian and English, and cultural nationalism. Although it lacks consistency and the use of Latin as the metalanguage restricted the target users to a limited group, it provides insights into nineteenth-century linguistics, lexicography and culture.

Chapter 7 focuses on John Jamieson's *Etymological Dictionary of the Scottish Language* (1808/1825), the first Scottish national dictionary. Having a reputation for its pioneering historical principles, the dictionary was of greater significance as a national and patriotic work, featuring widespread participation in the dictionary, coverage of cultural vocabulary, and presenting Scots as a distinct language. Though being criticized for excluding the lexis that Scots shared with English, it was regarded as a repository of national identity and cultural participation.

In Chapter 8, Wim Remysen and Nadine Vincent study Dunn's idea of a united and independent French Canada and the legitimacy of Canadian French. The topic is explored against the nineteenth-century sociopolitical context and the life experience of Dunn. They show that his vision was neatly reflected in the *1880 Glossaire franco-canadien,* which figured prominently in the development of Quebec lexicography.

In Chapter 9, Edward Finegan studies the 1828 edition of Noah Webster's *American Dictionary of the English Language*. He shows that the dictionary is a

manifestation of the lexicographer's nationalism and Christian commitments, through etymologies, definitions, and illustrative citations.

Chapter 10 explores Webster's unabridged dictionary of 1864. Peter Sokolowski focuses on the innovations of the revision work which set the course for the modernization of lexicography, featuring the organization of the editorial staff, the removal of Webster's etymologies, the historical ordering of the senses, the systematic pruning of the lexicon, the dual presentation of the engravings, the volunteer reading for citations, etc. The innovations, together with the keen business sense of the publisher, made the dictionary a great commercial success.

In Chapter 11, Ilya Vinitsky studies Vladimir Dahl's *Explanatory Dictionary of the Living Great-Russian Language* within the context of the Russian literature and culture and the religious and mystical doctrines of language. The dictionary reflected Dahl's view on language, that is, the word was a spiritual entity, a bridge between worlds, and a window onto the national soul. Viewing the Russian language as centered on the notion of family, he endeavored to reveal the communal spirit of Russia and create the national epic.

In Chapter 12, Mårten Söderblom Saarela discusses Banihûn and Pugong's *Manchu–Chinese Literary Ocean* (1821). The dictionary aimed at a greater integration of Manchu and Chinese and to serve the hybrid culture of lettered bannermen. He compares this dictionary with Callery's unfinished Chinese–French encyclopaedic dictionary, which was inspired by the same source dictionary, to show the importance of imperial Chinese lexicography. As the product of a specific historical period, the utility of the dictionary depended on the readers' proficiency in Manchu and their knowledge of the Confucian literary tradition.

In Chapter 13, Walter Hakala traces the century-long history of British lexicographic works on the Persian language, which concluded in 1892 with Francis Joseph Steingass's *A Persian–English Dictionary*. This chapter also documents the rise and fall of Persian as a transregional language. By the time Steingass produced his *Comprehensive Persian–English Dictionary* in 1892, the use of Persian in British India had been on the decline, which doomed the dictionary to failure.

Chapter 14 focuses on American missionary women lexicographers. Lindsay Rose Russell explores what made the missionary women lexicographers and what features their dictionaries contained. Compared with the large-scale dictionaries mentioned in the previous chapters, which aimed for national identities or featured refined lexicographic methods, the dictionaries made by the American missionary women are characterized as small-scale, pragmatic and appreciative of local languages and cultures.

In Chapter 15 Gabriella Safran discusses the historical confluence of Yiddish dictionaries and Jewish dialect comic books in the Russia Empire and the USA. Instead of assigning a social meaning to the dictionaries as seeking linguistic identity for a low-status and minority language, Safran focuses on the

popularization of Yiddish dictionaries and Jewish comic books and proposes that their success may be attributed to their ability to meet people's needs to produce low-status and high-status languages and perform comedy to impress others. Both were products of specific social conditions.

Chapter 16 is devoted to sign language dictionaries, which contributed to the recognition of the identity of the deaf community and the dissemination of sign languages. Jorge Bidarra and Tania Aparecida Martins start with a brief overview of the development of Libras dictionaries in Brazil from the nineteenth century to the twentieth century, showing that the development of sign languages was affected by educational philosophy and government policies. The authors then introduce the ongoing monolingual dictionary, PORLIBRAS, a project they have been working on.

Having different dictionaries discussed in one volume makes it possible and easy for the readers to compare and contrast. As shown, the authors of Chapters 5–9 share an interest in exploring the lexicographic works from the perspectives of language, culture, and nationalism. It also reveals that dictionaries produced in similar circumstances could yield different results. Steingass's *Comprehensive Persian–English Dictionary* failed, because of the limitation of Persian as a written elite language. In contrast, the dictionaries made by the American missionary women in Asia were successful since they insisted on the importance of local language users. These two examples also prove that the success of a dictionary depends in part on whether it could meet the needs of the users in a specific period, which is reinforced by the examples in Chapters 12 and 15. In addition, the dictionaries are explored from different perspectives. Some are focused on lexicographic innovations, such as Chapters 4, 5 and 10. Some, for example Chapters 6, 7, 8 and 9, attach greater weight to the social or cultural value of the dictionaries. Furthermore, the articles are not isolated, but related and compared. The contributors focus on different dictionaries. When they are exploring the dictionaries, it is usual that they compare the dictionaries with others, which also draws the readers' attention to the differences and similarities.

However, the book is not without shortcomings. It is not made clear by which principle the dictionaries were selected for study. Nearly half of the volume is devoted to English or English-related dictionaries. It would have better shown the panorama if more dictionaries of other languages were incorporated.

By offering an overview of nineteenth-century dictionaries and exploring dictionaries from varied perspectives, anyone interested in the history of lexicography would find this book informative, readable, and well researched.

References

Considine, J. 2019. *The Cambridge World History of Lexicography*. Cambridge: Cambridge University Press.
Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. London: Longman.

Dai Lingzhen
School of Foreign Studies
Minnan Normal University
Zhangzhou
P.R. China
(vivian0915@163.com)

Use of Dictionaries and Online Tools for Reading by Thai EFL Learners in a Naturalistic Setting

Atipat Boonmoh, *School of Liberal Arts,*
King Mongkut's University of Technology Thonburi,
Bangmod, Thailand (atipat.boo@kmutt.ac.th)

Abstract: Dictionaries and online tools are regarded as important tools for finding out the meanings of unknown words or checking the usage of interesting words. This study investigated students' use of dictionaries and online tools in a natural setting by identifying the type of words they look up in their dictionaries and the types of dictionaries they use for the look-up. Fourteen learners from a Thai public university participated as the subjects by filling in a task record form which included the names of the dictionaries they were using, look-up words, meanings selected for the words, and their parts of speech. The look-up words were analyzed to find out the frequencies of words, parts of speech, and meanings by checking with an advanced learner dictionary. The findings show that the subjects mainly selected words that were nouns, verbs, and adjectives. They tended to choose "high frequency" for the lookups. For their convenience, subjects used either an online translation tool, 'Google Translate', or bilingual dictionaries through mobile devices. The results may help us understand EFL dictionary users' behaviors in using dictionaries and provide interesting implications for language teachers.

Keywords: DICTIONARY USE, DICTIONARY LOOKUPS, THAI EFL LEARNERS, GOOGLE TRANSLATE, DICTIONARY RECORD FORM, ONLINE DICTIONARIES

Opsomming: Thaise EVT-leerders se gebruik van woordeboeke en aanlyn hulpmiddels vir lees in 'n natuurlike omgewing. Woordeboeke en aanlyn hulpmiddels word as belangrike werktuie beskou om die betekenisse van onbekende woorde te bepaal of om die gebruik van interessante woorde na te gaan. In hierdie studie is studente se gebruik van woordeboeke en aanlyn hulpmiddels in 'n natuurlike omgewing bestudeer deur die tipe woorde wat deur hulle nageslaan word in hul woordeboeke asook die tipe woordeboeke wat hulle vir die naslaanproses gebruik, te identifiseer. Veertien leerders van 'n Thaise staatsuniversiteit het as proefpersone deelgeneem deur 'n taakkontrolevorm te voltooi wat die name van die woordeboeke wat hulle gebruik het, die woorde wat nageslaan is, die betekenisse wat vir die woorde geselekteer is, en hul woordsoorte, ingesluit het. Die woorde wat nageslaan is, is geanaliseer om die frekwensies van woorde, hul woordsoorte, en betekenisse te bepaal deur 'n gevorderde aanleerderswoordeboek te raadpleeg. Die bevindings dui daarop dat die proefpersone hoofsaaklik substantiewe, werkwoorde en adjektiewe geselekteer het. Hulle was geneig om voorkeur aan "hoëfrekwensie" in

die naslaanproses te verleen. Geriefshalwe het die proefpersone óf 'n aanlyn vertalingshulpmiddel, 'Google Translate', óf tweetalige woordeboeke met behulp van mobiele toestelle gebruik. Die resultate mag ons dalk help om EVT-gebuikers se gedrag in die gebruik van woordeboeke te verstaan en lewer interessante implikasies vir taalonderwysers.

Sleutelwoorde: WOORDEBOEKGEBRUIK, WOORDEBOEKNASLAANPOGINGS, THAI EVT-LEERDERS, GOOGLE TRANSLATE, WOORDEBOEKKONTROLEVORM, AANLYN WOORDEBOEKE

Introduction

A dictionary is an essential tool when learning a foreign language. Learners consult their dictionaries for many purposes for different tasks. In a reading task, learners are likely to consult their dictionaries to find out the meaning of unknown or unfamiliar words. The main focus for dictionary use for reading as suggested by Schofield (1982) is to find out the specific meaning that is relevant to the context of the unfamiliar word or phrase, rather than to find out its spelling or grammatical characteristics. Learners can find the information about the meaning by using either a monolingual learner's dictionary or a bilingual (L2>L1) dictionary.

In recent years, it has been shown that new technology is being integrated into our lives at a great rate. The widespread use of mobile devices, especially computers and smartphones, significantly changes the ways of learning in many contexts, including language learning (Kukulska-Hulme 2009). This also changes the ways students consult their dictionaries, and internet or online dictionaries have gained popularity among EFL students (Dashtestani 2015). A number of studies (e.g., Fallianda 2020, Alhatmi 2019, Metruk 2017, Aslan 2016, Anongchanya and Boonmoh 2015, Ding 2015) have shown a growing number of online dictionary users in many countries around the world.

In Thailand, the number of smartphone users have increased rapidly. Studies into dictionary use (e.g. Boonmoh 2018, Chotkarnchanawat et al. 2017, Inpin 2016) show that students reported using dictionaries and online dictionaries both in classrooms and at home. Although online dictionaries are common in Thailand, they tend to be ignored by language teachers. Few studies that investigated dictionaries and online tools were conducted in classroom settings where their look up behavior may be different from natural settings.

Dictionary use is a private activity and the learners are often inclined to be secretive (Nesi and Boonmoh 2009), especially when they are of low L2 proficiency and when they feel that their teachers disapprove of the types of dictionaries they use. The choice of dictionary type also depends on individual learners. A number of previous studies (Liang and Xu 2018, Ding 2015, Dziemianko 2010) into dictionary use have been conducted in controlled experimental settings i.e., controlling the use of a specific type of dictionaries, the type of reading

passages, the controlled setting of the experiment and the time spent. However, the findings may not be applicable to a real context where learners can choose which texts to read and to freely consult any dictionaries. This study, therefore, aims to find out which dictionaries students use and what type of words students look up in dictionaries when they are reading passages in their own free time.

Literature Review: Uncovering dictionary use

There are different methods to investigate dictionary use and dictionary look-up behavior. The following are a few but detailed reviews of previous studies with special reference to the methods used, which provide justification for the choice of this study's methodology.

One of the most common methods to research dictionary use is perhaps questionnaire research. Many studies in dictionary research (Fallianda 2020, Hojatpanah and Dashtestani 2020, Alhatmi 2019, Abbasi et al. 2019, Chotkarn-chanawat et al. 2017, Aslan 2016, Boonmoh 2010) have employed this method. Although questionnaire survey research can be used as a way of obtaining results from a great number of respondents and has the potential for generating and testing a hypothesis because the large number of respondents can also mean better representation of the whole population, one basic problem associated with this type of survey as discussed by Nesi (2000) is the accuracy of responses. It is difficult for subjects to give correct information for questions which require them to recall their previous dictionary use behaviors. Some questions may require a capacity to recall, retrospect and comprehend beyond the abilities of the average dictionary user. Crystal (1986: 78) points out that retrospective questionnaire items place high demands on subjects' memories. He even challenges if anyone could "*confidently write down when they last used a dictionary, why they used it, and how often they consult one*". For this reason, it may not be a good idea to use questionnaires to investigate dictionary use behavior.

Interviews can be used to elicit opinions, and interactive settings are another step towards gathering more direct evidence of dictionary look-up behavior. Previous research which employs interviews together with other research instruments includes Hojatpanah and Dashtestani 2020, Baskin and Mumcu 2018, and Dashtestani 2013). The interview questions can be structured but can also be flexible in the sense that interviewers may ask further questions related to the interviewees' replies. It has been suggested, however, (e.g., Gorden 1980) that the more structured the interview, the more interviewees will feel like '*subjects*' rather than '*informants*'. Making the interviewees feel like '*subjects*' might affect how they supply information as they might try to please researchers by providing information that they think we want to hear. Another basic problem when using an interview as a research instrument is similar to a problem using a questionnaire. The responses may not be accurate. Hatherall

(1984: 184) questioned the success of questionnaire data as a means of recording subjects' behavioral acts: "*Are subjects here saying what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three?*

Observation can be an effective way to describe dictionary use behavior as Hatherall (1984) recommends that watching dictionary users in action is "*the only reliable method of collecting data on dictionary user behavior*". It is, however, very difficult for teachers or researchers to see what is happening during a student's dictionary consultation as the dictionary consultation is a private activity. A less obstructive method to observe how dictionaries are used can be done by using 'log file' on computers. '*Log file*' is a file that lists actions that have occurred (through keystroke logging). With log file analysis tools, it is possible to record all information about the use of a computer: what users key in, what they delete, what words in a dictionary they look up, etc. Studies that have employed log files include, for example, Koplenig et al. (2014), Chen (2011), Laufer and Levitzky-Aviad (2006), and Hulstijn and Trompetter (1998). Some limitations of using 'log file' as a research instrument listed by Lew (2015) are that "*log files will rarely tell us what the context of dictionary use is: what activity the user is involved in, what particular problem they are trying to solve, and the level of success and satisfaction achieved in the consultation*". Since this study aims at looking at dictionary use in a naturalistic setting, learners will not be restricted to using only dictionaries on computers but may also access dictionaries on their own devices and/or in printed forms.

Another method of investigating dictionary use involves lookup record forms. Some studies that have employed this method are Atkins and Varantola (1998), Diab and Hamdan (1999), Al-Ajmi (2002), and Frankenberg-Garcia (2005). These studies were conducted in a natural setting, reducing the likelihood of producing distorted data. Using dictionary record form, each subject could do the work at their own pace and they could use their dictionaries to solve their reading problems without being observed. However, asking students to record words they look up may not be appropriate for research into how dictionaries are really used. This is because the focus of the study (by using lookup record forms) would be on the final decision of the students rather than on the entire lookup process.

More studies using more naturalistic approach are needed to explore how learners use dictionaries in real situations, particularly in language learning contexts such as reading, writing and translation (Liang and Xu 2018), as dictionary consultation is a private matter and there is no way of discovering what people actually do when they use a dictionary without interfering with their natural behavior. Using a questionnaire would require too much memory recall from the students. Using a log file can help reveal this, but it is limited to observing dictionary use on mobile devices. When learners access their dictionaries or reference tools in printed format, log files cannot be used. Observation can look at how students use dictionaries in a natural setting but the fact that the students are being observed may affect their dictionary use behaviors.

Asking the students to verbalize would inevitably disrupt the subjects' working processes. Using a dictionary record form can reveal only the final production of dictionary use i.e., what words they looked up for their reading. Taking all the type of research methods into consideration, the most appropriate research method for this study is, therefore, the use of a dictionary record form since the purpose of this study is purely to investigate the type of dictionaries and the type of words the students look up in dictionaries when they are reading passages at a time and place of their choice, and they can read and look up words at their own speed without any disruption.

Purposes of the study

The purposes of this study are to find out the following.

(a)   What types of words were selected?
(b)   From the selected words, what frequency level do they belong to?
(c)   What types of dictionaries did students use?

Methodology

Participants

The participants of this study were 14 learners who enrolled in an English course entitled "*Reading and Vocabulary Level 2*" in the academic year 2020. These 14 learners were support staff members at a Thai public university in Bangkok, Thailand. Under the university Training Roadmap, six English courses were available for the staff: *Reading and Vocabulary Level 1 and Level 2*, *Listening and Speaking Level 1 and Level 2*, and *Writing and Grammar Level 1 and Level 2*. The 14 participants were from different departments/divisions of the university. Their ages ranged from 34 years old to 53 years old. Of these, 12 had at least a bachelor's degree and the remaining two participants had vocational diplomas. Their English language proficiency was considered low. These participants were chosen for two reasons. First, they enrolled in and attended the course voluntarily. Second, the content of the course was about strategies in learning vocabulary and strategies in reading. The course aimed to develop vocabulary knowledge through reading and also involved reading strategies, dealing with unknown words, and using dictionaries.

Data collection

The reading text

As a part of the course requirement, the students needed to complete a sum-

mary task where they had to select a news article in English, read it thoroughly at their own pace and at a time and place of their choice, and write a summary in Thai. The news articles were chosen because they could lead students to many benefits such as increasing their awareness and motivation and encouraging them to bring a sense of reality and authenticity to the classroom (Rao 2019).

Ten news articles were taken and adapted from a BBC website and students were given choices to choose one of these ten articles which included a variety of news such as business, education, entertainment, environment, science and technology, and national and international issues:

— Air pollution: Thailand schools still closed due to 'unhealthy' smog levels (247 words)
— Could your firm move to a four-day week? (389 words)
— EU-Swiss share trading row: What does it mean? (325 words)
— Fukushima nuclear disaster: Abandoned town allows first residents home (215 words)
— Jack Ma defends the 'blessing' of a 12-hour working day (338 words)
— Salmon farming 'pays £100m in British taxes' (392 words)
— Starbucks to pay staff tuition fees (371 words)
— Tesla to raise prices and keep more stores open (394 words)
— TikTok app banned by US Army on work mobile phones (354 words)
— Where we are with Brexit — in 300 words (355 words)

Dictionaries and online tools

At the beginning of the course, the students were introduced to different types of dictionaries and online tools that are useful for reading. For the summary task, the students were instructed to use any dictionary of their choice i.e., bilingual or learner's dictionaries, online dictionaries from computers, or dictionaries or applications on mobile devices.

Task record form

After the students chose the news articles, they were given one week to complete the summary task. They were instructed to do this task at their own pace and at a time and place of their choice. The students had to submit two documents: the summary, and the task record form which included information about (1) their reason for choosing the news article, (2) looked-up words, their meanings, and word classes, and (3) the type and the specific name of the dictionary used. (See Appendix A for the task record form.)

Data analysis

The researcher studied the elicited data and tabulated them accordingly i.e., looked-up words, their word classes, and their word frequency were put in a spreadsheet. All the looked-up words were analyzed to see whether the words fell in High, Mid, or Low-frequency words based on Longman Communication 9000 (LC9000) in Longman Dictionary of Contemporary English Online 6th edition (LDOCE), which is based on 390 million words in the Longman Corpus Network. This dictionary was chosen because of two reasons. First, it contains features typically found in major advanced learners' dictionaries. Second, this dictionary is well-known in Thailand.

LDOCE contains words appearing in the Academic Word List (AWL). These are important words to know when reading and writing academic assignments. It also provides the Longman 9000 keywords (LC9000) which are the most important 9,000 words to learn in English. The words can be classified into three levels: High frequency words, Mid frequency words, and Low frequency words. The three red dots indicate the top 3000 words, the two red dots indicate the next most important words, and the one red dot indicates the less frequent yet important next 3000 words.

However, in this dictionary there are some other words which do not belong to any of the 3 categories. These words occur less frequently than others, but they are still important and deserve an entry in the dictionary.

From the raw data, the reasons for choosing the news articles were noted, and the total number of looked-up words was counted and categorized into two themes: word classes and word frequency. The names of dictionaries were counted and classified. (See Appendix B for full details of all lookups.) The data from students' written summaries were not used in the analysis.

Findings

It can be seen from Table 1 that only four (out of ten) news articles were chosen by the students. P1 to P4 chose to read the article about air pollution in Thailand and P5 to P8 chose to read about a nuclear disaster in Fukushima. The next three students (P9 to P11) chose to read about an American multinational chain of coffeehouses called Starbucks and the remaining three students (P12 to P14) chose to read about a video-sharing social networking service called TikTok. When asked about the reasons, common themes emerging from this were having some background knowledge of the topic or having familiarity with the topic, having interest in the topic, and the short length of the news article.

Table 1: News article read by students, their reasons for choosing it and number of dictionary lookups

Students	News Article Read	Reasons	Number of Lookups
P1	Air pollution: Thailand schools still closed due to 'unhealthy' smog levels	*I think it will not be too difficult.*	15
P2		*It's about Thailand.*	11
P3		*I have some background knowledge.*	8
P4		*Familiar content so it will be easy to understand.*	13
P5	Fukushima nuclear disaster: Abandoned town allows first residents home	*I have some background knowledge.*	15
P6		*It is the shortest. It seemed to be the easiest.*	11
P7		*It's short.*	9
P8		*I want to keep updated about it.*	10
P9	Starbucks to pay staff tuition fees	*I like Starbucks Coffee.*	10
P10		*I know about Starbucks.*	11
P11		*Interesting*	10
P12	TikTok app banned by US Army on work mobile phones	*My daughter uses it and I want to know why it is banned.*	17
P13		*I heard about it and want to know more.*	12
P14		*My children use it.*	13
Total number of lookups 95 (Noun) + 44 (Verb) + 22 (Adjective) + 4 (Adverb)			165

In terms of dictionary lookups, the total number of consultations completed by all the 14 subjects were 165. The number of lookups completed by each subject ranged from 8 to 17. The average lookups were 12. Of these 165, 95 (57.6%) lookups were words classified by LDOCE as nouns, 44 (26.7%) lookups were classified as verbs, and 22 (13.3%) lookups were classified as adjectives. Adverbs were among the least common words that the subjects looked up in the dictionaries. These only accounted for 4 (2.4%) lookups which were "*currently, recently, increasingly*", and "*partially*". It should be noted that the total number of lookups does not equate the total number of words. This is because some words were looked up by several students. For example, the word 'plant' was looked up by 4 students, the word 'Chernobyl' was looked up by 3 students, the word 'decontamination' was looked up by 2 students. There were altogether 9 lookups, but these lookups are counted as three words. (See Appendix B for full details of all lookups and words.)

Since the subjects chose the article of their choice to read, there are certainly words that they did not know the meaning of or words that could help them with their comprehension. The words that they were likely to look up might be content words and specific words that they need to know about the topic being read. For example, "*EU, Arizona, Brexit, Pret a Manger*" were looked up by the subjects who read the article "*Starbucks to pay staff tuition fees*". It is obvious that most of the lookups were content words i.e., nouns, verbs, and adjectives, because these words carry meanings and must be included in a sentence for it to make sense, while the others were function words which only add proper grammatical structure and flow to the sentence. This might explain why the subjects looked up more content words and specific words than function words.

Table 2:   Lookups classified by frequencies

	Frequency	Word class	Lookups	Total lookup
Longman Communication 9000	High frequency	N.	29	52 (31.5%)
		V.	16	
		Adj.	5	
		Adv.	2	
	Mid frequency	N.	26	55 (33.3%)
		V.	20	
		Adj.	8	
		Adv.	1	
	Low frequency	N.	19	23 (14%)
		V.	2	
		Adj.	1	
		Adv.	1	
	Unidentified	N.	21	35 (21.2%)
		V.	6	
		Adj.	8	

From Table 2, upon examining the subjects' frequency of lookup words, it can be seen almost 80% of all the lookups (130 lookups) were words that are listed in the LC9000. Of these, 52 lookups (31.5%) belong to the High-frequency category, and 55 lookups (33.3%) belong to the Mid-frequency category, and 23 lookups (14%) belong to the Low-frequency category.

Only 35 lookups were classified as unidentified because the frequency marks are not shown in the Longman Dictionary of Contemporary English. These are, for example, "*Chernobyl, Arkansas, Arizona, Brexit, EU, Pret A Manger, Newsroom, tsunami, haze, incense*" for nouns, "*Hosing, expediting, lip-synching*" for verbs, and "*grainier, toxic, quirky*" for adjectives.

Table 3:   Examples of Lookups classified by Longman Communication 9000

Participants	Titles of news articles	Lookups listed in LC9000	Unidentified	
				Proper noun
P1 P2 P3 P4	Air pollution: Thailand schools still closed due to 'unhealthy' smog levels	effect, cases, lung, smog, exhaust, blame, reduce, tackle, tiny, unhealthy, harmful	Haze, incense, hosing, grainier, toxic	—
P5 P6 P7 P8	Fukushima nuclear disaster: Abandoned town allows first residents home	plant, inhabitants, recovery, radiation, infrastructure, accuse, flee, vast	tsunami, expediting	Chernobyl
P9 P10 P11	Starbucks to pay staff tuition fees	employee, tuition, incentive, under-graduate, enroll, offer, obtain, financial, available	—	Brexit, EU, Pret A Manger, Arizona
P12 P13 P14	TikTok app banned by US Army on work mobile phones	cyber threat, policies, privacy, authorities, scrutiny, hire, restrict, issue, wary	lip-synching, quirky	Arkansas

Table 3 shows some words that were searched by the students in order to complete the summary task. It is clear most searched words are content words and are listed in the LC9000 (Also see Appendix B). These searched words are important for the students as they help the students to comprehend the news articles effectively.

For example, Participant 10 looked up these 11 words: "*incentive, initial, expand, obtain, Brexit, Pret A Manger, firm, offer, undergraduate, financial, reduce*". These are key words that help Participant 10 understand the main idea of the news article 'Starbucks Coffee <u>offers</u> an <u>incentive</u> to its staff in UK who want to <u>obtain</u> an <u>undergraduate</u> degree by giving <u>financial</u> support to them'.

Taking the titles of articles into consideration, it is obvious that many of the lookups that are classified as unidentified are proper nouns or words that are specially related to the specific topics the participants chose to read. For example, the words "*haze, incense, hosing, and toxic*" were searched by students as these words are related to the 'toxic <u>haze</u> that can be reduced by <u>hosing</u> down the streets and not burning <u>incense</u>'. The proper nouns "*EU, Arizona, Brexit, Pret a Manger*" were looked up as it specifically discussed about 'recruitment problems Pret a Manger may encounter after Brexit from the EU'

Although this study did not take the success of dictionary consultation and the correctness and overall quality of the written summaries into account, it should be noted that some subjects misunderstood the news articles. Part of

this misunderstanding was due to their dictionary lookups. For example, Participant 6 looked up the word 'plant' from *Google Translate* and wrote the meaning "ปลูก" /to plant/ in the dictionary record sheet. His written summary talks about planting trees nearby Fukushima power plant.

Table 4:   Dictionaries and online tools consulted by students

Dictionaries and Online Tools	Participants who used this type														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Google Translate	/	/	/	/	/	/	/	/	/	/	/	/	/	/	14
www.google.com										/	/	/			3
Longdo Dict					/										1
Sanook Dict								/							1
Line Dict						/									1
LDOCE				/											1
Se-Ed's Modern English-Thai dictionary	/		/												2
Total	2	1	2	1	2	1	2	1	2	2	2	2	1	1	23

Table 4 shows the proportion of dictionaries and online tools that the 14 students consulted while reading the news articles. *Google Translate* is by far the most popular online translation tool that they used for the summary task. All 14 students used *Google Translate* from their mobile devices. It can be seen that not everyone uses only one dictionary. All 14 students used *Google Translate* as a base dictionary and some students used other dictionaries.

Three students used websites such as www.google.com to search for unknown words. For example, in the task record form, Participant 10 wrote that he looked up the word "Pret A Manger" from *Google Translate*, but the translation was "แกล้งทำเป็นรางหญ้า" /pretend to be a manger/. He then looked up "Pret A Manger" from www.google.com. and later understood that it is a name of a well-known sandwich company. Interestingly, Participant 12 reported in the task record form about searching for the words "*military, army, navy, Arkansas*" from *Google Images*.

Three students used online English–Thai dictionaries from their mobile devices. Online English–English dictionaries were the least popular as can be seen that only one participant reported using Longman Dictionary of Contemporary English Online. This may be because LDOCE was used as an example of a learner's dictionary introduced to the students at the beginning of the course. It is also interesting to note that taking the mode of dictionaries into consideration, online-based dictionaries outnumber paper-based dictionaries by 18 to 2. Only two students used an English–Thai dictionary in book format for this summary task.

Discussion and Conclusion

This study reveals insights into how the students selected a news article of their own choice to read at their own time, the type of words they looked up from their dictionaries and online tools as well as their choice of those tools used to complete the summary task in a naturalistic setting.

The findings of this study showed that when choice and adequate time were given, students chose to read the news articles based on their familiarity or the background knowledge they had on the topic, their interests, and the perceived difficulty of the text. This finding is obvious, but it can provide some implications for research design of receptive use of dictionary use research. In order to better understand students' dictionary use behavior, it is important that they can choose reading texts according to their own interests or they can have some degree of freedom to choose reading texts. They would be more motivated to engage with the text (Butcher and Kintsch 2003), and the more background knowledge they have on the text, the more likely they will be able to make sense of what is being read (Pardo 2004).

In the analysis of word types and their frequency, it is found that the students looked up content words (noun, verb, adjective: 97.6%) much more than function words (adverbs: 2.4%). This finding is not surprising because in any given text it is typical that the number of content words is higher than function words. The fact that the subjects looked up fewer adverbs might be because there were fewer of these than other kinds of content word in the text. On the other hand, content words are words that are related to both meaning and comprehension, and acquisition of content words is essential for successful second language use. Content words are usually polysemous (Crossley et al. 2010), they carry multiple meanings which can cause problems for students in meaning comprehension. This may explain why content words were looked up much more than function words. This finding is consistent with that of Liang and Xu (2018) who reported Chinese learners indicated their tendency to look up content words more than function words.

The study also showed that the students consulted high-frequency words most often as can be seen that 80% (130 out of 165) of all lookups are listed in the LC9000. This finding corresponds to that of Liang and Xu (2018) and Koplenig et al. (2014). The high proportion of high-frequency words that were looked up may reflect the fact that students' vocabulary was limited and they relied heavily on dictionaries and online tools to comprehend their meaning.

In terms of the students' choice of dictionaries and online tools, analysis of the data indicated that the students used bilingual dictionaries more frequently than monolingual ones as well as used online dictionaries more than paper-based dictionaries. This accords with a survey finding conducted by Fallianda (2020) that pre-intermediate and intermediate students use bilingual dictionaries more often than monolingual dictionaries. The finding of this study is also consistent with those in Tong (2019), Małgorzata (2016), and Dashtestani (2013) in

that bilingual and online dictionaries were the most popular among the students.

This study found that an overwhelming number of students reported that they used *Google Translate* to do the summary task. This finding is similar to Małgorzata's investigation (2016) that the students highlighted their preference for online dictionaries and online resources rather than traditional paper dictionaries. However, it is interesting to note that *'Google Translate'* is not a dictionary. It is an online translating tool, but it has gained much popularity from many users from which can be inferred that the students found it convenient to find the meaning of words. Even though teachers encourage students to use learner's dictionaries (Boonmoh 2010), the students seemed to prefer to use other online dictionaries and tools when they have access to their own mobile devices.

Interestingly, the finding of the current study is contrary to that of Ding (2015) and Diab and Hamdam (1990) who found that monolingual dictionaries were preferred by their students. This may be explained by the fact that the students of these two studies were English major students who presumably were advanced learners of English, and they used dictionaries in their translation classes. Another explanation is related purpose of dictionary use. Using a dictionary for translation task and using a dictionary for a summary may affect the choice of dictionary use. Since the students in the current study were lower intermediate learners of English and they preferred bilingual dictionaries, this may suggest that they are not confident enough about their English and they might feel more comfortable using bilingual dictionaries than English–English dictionaries as they could understand the meanings of the target words right away. This can be further supported by Corrius and Pujol (2010) in that bilingual dictionaries generate confidence among users.

Conclusion

This study investigated the use of dictionaries and online tools by students in a naturalistic setting by identifying students' justification when selecting a news article to read in their own time and exploring the frequency and type of words looked up in dictionaries and online tools by the students. It also identified dictionaries and online tools used by the students.

Based on the findings, it can be proposed that when teaching students in a reading class, teachers should select reading passages that are authentic and relevant to students' background knowledge. Students should be allowed an opportunity to choose to read a text of their interest and at their own time. As acquisition of content words is essential for successful second language use, teachers may consider providing explicit training in dictionary use. Teacher may introduce the students to different learners' dictionaries and provide information on the criteria according to which dictionaries determine the priority in vocabulary and word frequency (Metruk 2017). For example, the

LC9000 features in LDOCE can help students be well-informed in vocabulary learning and use. This will also help students know which words are assigned priority. Since words in high- and mid-frequency categories were looked up the most, this may suggest students' limited vocabulary knowledge. Therefore, teachers should not ignore vocabulary teaching and reading strategies so that the students can increase their reading proficiency.

One last important result emerging from this study is the frequent use of '*Google Translate*'. Since this study was conducted in a natural setting and no particular dictionaries or online tools were imposed on the students, it is clear that the use of '*Google Translate*' will become a fact of life for EFL learners in Thailand and other countries. This increase in "*Google Translate*' use will eventually affect classroom environments. The presence of "*Google Translate*" will continue, as Stirling (2005: 64) suggested "they seem to be here to stay, just as calculators, once welcomed by students and rejected by teachers, have stayed."

The decision whether to ignore "*Google Translate*" use in classroom or to help students make full use of "*Google Translate*" will depend on the teachers. Language teachers should explore this thoroughly in terms of advantages and disadvantages so that they can be in a position to train students to use "*Google Translate*" or to integrate Google Translate along with other online dictionaries effectively. For example, when using "Google Translate" to find the meaning of the words, the teachers should make the students beware that when using a dictionary or online tool to find the meaning of the words, they need to select the appropriate entries for the context so that they can better understand the text. Teacher may suggest alternative ways to look for meanings of the unknown words such as using *Google Images*.

This study provides insight into dictionaries and online tools use in natural setting. However, it has some limitations. The number of students in the study is not high. The meanings that the students supplied in the task record form after look-up were not taken into account although the finding is based on the task record form. Although the students were allowed to choose one of the ten news articles, the students were not entirely free to choose what they wanted to read. This might have provided a better picture of how the students use dictionaries and online tools to meet their reading needs. To improve these, future studies with a larger population, with free choice of text read and with mixed research methods should be conducted. Studies into success in dictionary consultation taking into account task record form should be conducted.

References

Abbasi, W.T., M.M. Ahmad, and F.A.E. Mohammed. 2019. Learners' Perceptions of Monolingual Dictionaries in Learning English as a Foreign Language. *International Journal of Education & Literacy Studies* 7(3): 10-18.

Al-Ajmi, H. 2002. Which Microstructural Features of Bilingual Dictionaries Affect Users' Look-up Performance? *International Journal of Lexicography* 15(2): 119-131.

Alhatmi, S. 2019. A Survey Study of the Dictionary Use Sub-strategies of English Majors in Saudi Arabia: Dictionary Related Aspects. *English Language Teaching* 12(3): 139-160.

Anongchanya, E. and A. Boonmoh. 2015. The Effectiveness of the Use of Dictionary Applications in Smartphones in Reading an English-language Passage and Writing a Summary in Thai. *Proceedings from ICLLCS 2015, 2nd International Conference on Language, Literature, and Cultural Studies, 20-21 August 2015, Chonburi, Thailand:* 47-58. Saen Suk, Thailand: Department of Western Languages, Faculty of Humanities and Social Sciences, Burapha University.

Aslan, E. 2016. A Study on the Use of Mobile Dictionaries in Vocabulary Teaching. *Journal of Language and Linguistic Studies* 12(1): 1-8.

Atkins, B.T.S. and K. Varantola. 1998. Monitoring Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators:* 83-122. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.

Baskin, S. and M. Mumcu. 2018. Dictionary Culture of University Students Learning English as a Foreign Language in Turkey. *International Education Studies* 11(3): 101-114. Doi: 10.5539/ies. v11n3p101.

Boonmoh, A. 2010. Teachers' Use and Knowledge of Electronic Dictionaries. *ABAC Journal* 30(3): 56-74.

Boonmoh, A. 2018. The Use of Online Dictionaries for Writing by Thai University Students. *Proceedings from ICES 2018, the 1st International Conference on English Studies: "Innovation in English Language Teaching and Learning", Krabi, Thailand, 28–29 June 2018:* 333-340. Muang, Songkhla: Department of Western Languages, Faculty of Humanities and Social Sciences, Thaksin University.

Butcher, K.R. and W. Kintsch. 2003. Text Comprehension and Discourse Processing. Healy, A.F. and R.W. Proctor. 2003. *Handbook of Psychology: Experimental Psychology. Volume 4:* 575-595. New York: John Wiley.

Chen, Y.Z. 2011. The Use of Bilingualized English–Chinese Learner's Dictionaries: A Survey and An Experiment. *Lexicographical Studies* 2: 141-158.

Chotkarnchanawat, R., C. Liangpanit and S. Wasuntarasophit. 2017. Graduate Students' Reflections on the Use of Online Dictionaries for Writing a Thesis. *Journal of Humanities and Social Sciences* 13(2): 223-246.

Corrius, M. and D. Pujol. 2010. A Bilingualised English Dictionary for Catalan Speakers. *Lexikos* 20: 109-122.

Crossley, S., T. Salsbury and D. McNamara. 2010. The Development of Polysemy and Frequency Use in English Second Language Speakers. *Language Learning* 60(3): 573-605.

Crystal, D. 1986. The Ideal Dictionary, Lexicographer and User. Ilson, R. (Ed.). 1986. *Lexicography: An Emerging International Profession:* 72-81. Manchester: University Press.

Dashtestani, R. 2013. EFL Teachers' and Students' Perspectives on the Use of Electronic Dictionaries for Learning English. *CALL-EJ* 14(2): 51-65.

Dashtestani, R. 2015. Moving Bravely Towards Mobile Learning: Iranian Students' Use of Mobile Devices for Learning English as a Foreign Language. *Computer Assisted Language Learning* 29(4): 815-832.

Diab, T.A. and J.M. Hamdan. 1999. Interacting with Words and Dictionaries: The Case of Jordan EFL Learners. *International Journal of Lexicography* 12(4): 281-305.

Ding, J. 2015. A Study of English Majors in a Chinese University as Dictionary Users. *Lexicography: Journal of Asialex* 2(1): 5-34.

Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

Fallianda. 2020. A Survey of Indonesian Students' Use of Dictionaries. *Lexikos* 30: 609-628.

Frankenberg-Garcia, A. 2005. A Peek into What Today's Language Learners as Researchers Actually Do. *International Journal of Lexicography* 18(3): 335-355.

Gorden, R.L. 1980. *Interviewing: Strategy, Techniques, and Tactics*. Third edition. Homewood: Irwin-Dorsey.

Hatherall, G. 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983:* 183-189. Tübingen: Max Niemeyer.

Inpin, B. 2016. Smartphones: Noteworthy Tools for Enhancing EFL Students' English Language Learning. *Academic Journal: Uttaradit Rajabhat University* 11(1): 167-180.

Hojatpanah, S. and R. Dashtestani. 2020. Electronic Dictionaries as Language Learning Tools for Iranian Junior High School Students. *CALL-EJ Online* 21(1): 79-96.

Hulstijn, J.H. and P. Trompetter. 1998. Incidental Learning of Second Language Vocabulary in Computer-assisted Reading and Writing Tasks. Albrechtsen, D., B. Henriksen, I.M. Mees and E. Poulsen (Eds.). 1998. *Perspectives on Foreign and Second Language Pedagogy:* 191-200. Odense, Denmark: Odense University Press.

Koplenig A., P. Meyer and C. Müller-Spitzer. 2014. Dictionary Users Do Look Up Frequent Words. A Log File Analysis. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 229-250. Lexicographica Series Maior 145. Berlin/Boston: De Gruyter.

Kukulska-Hulme, A. 2009. Will Mobile Learning Change Language Learning. *ReCALL* 21(2): 157-165.

Laufer, B. and T. Levitzky-Aviad. 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' — A Dictionary for Production in a Foreign Language. *International Journal of Lexicography* 19(2): 135-155.

Lew, R. 2015. Opportunities and Limitations of User Studies. *OPAL — Online publizierte Arbeiten zur Linguistik* 2015(2): 6-16.

Liang, P. and D. Xu. 2018. An Empirical Study of EFL Learners' Dictionary Use in Chinese–English Translation. *Lexikos* 28: 221-244.

Małgorzata, K. 2016. Dictionary-using Skills of Translation Students. *The Central European Journal of Social Sciences and Humanities* 4: 235-242.

Metruk, R. 2017. The Use of Electronic Dictionaries for Pronunciation Practice by University EFL Students. *The Journal of Teaching English with Technology* 17(4): 38-51.

Nesi, H. 2000. *The Use and Abuse of EFL Dictionaries: How Learners of English as a Foreign Language Read and Interpret Dictionary Entries.* Lexicographica Series Maior 98. Tübingen: Max Niemeyer.

Nesi, H. and A. Boonmoh. 2009. A Close Look at the Use of Pocket Electronic Dictionaries for Receptive and Productive Purposes. Fitzpatrick, Tess and Andy Barfield (Eds.). 2009. *Lexical Processing in Second Language Learners:* 67-81. Clevedon, UK: Multilingual Matters.

Pardo, L.S. 2004. What Every Teacher Needs to Know about Comprehension. *The Reading Teacher* 58(3): 272-280.

Rao, P.S. 2019. The Impact of Using Newspapers on Teaching Language Skills to the Second Language Learners of English. *ELT Vibes: International E-Journal For Research in ELT* 5(2): 154-170.

Schofield, P. 1982. The Role of Bilingual Dictionaries in ESL/EFL: A Positive View. *Guidelines* 4(1): 84-98.

Stirling, J. 2005. The Portable Electronic Dictionary — Faithful Friend or Faceless Foe? *Modern English Teacher* 14(3): 64-72.

Tong, L.T.M. 2019. Vietnamese EFL Students' Habits of Using Dictionaries in Translation Practice in ESP Courses. *Journal of English Language Teaching and Linguistics* 4(3): 305-318.

Appendix A: Dictionary Record form

1.   Which news article did you choose to read?

　　　_____
　　　_____
　　　_____

2.   Why did you choose this article?

　　　_____
　　　_____
　　　_____

3.   What are the most interesting words that you learned? What words did
　　　you look up?

Words	Part of speech	Meaning(s)

4.   Please write down the names of dictionaries and online tools that you
　　　used in this task.

　　　(1)   …………………………………………………………………..
　　　(2)   …………………………………………………………………..
　　　(3)   …………………………………………………………………..

_____

Appendix B: Details of Lookups and words

No.	Words	Classified by LC9000	Searched by	Frequency	No.	Words	Classified by LC9000	Searched by	Frequency
				NOUN					
1	plant	High	P5 P6 P7 P8	4	31	lungs	Medium	P1	1
2	army	High	P12 P14	2	32	navy	Medium	P12	1
3	authorities	High	P5 P7	2	33	personnel	Medium	P14	1
4	concern	High	P7 P13	2	34	privacy	Medium	P14	1
5	military	High	P12 P13	2	35	residents	Medium	P6	1
6	policies	High	P12 P13	2	36	Spokeswoman	Medium	P11	1
7	quality	High	P2 P4	2	37	strain	Medium	P11	1
8	threat	High	P12 P13	2	38	tuition	Medium	P9	1
9	attempt	High	P8	1	39	undergraduate	Medium	P10	1
10	billion	High	P14	1	40	infrastructure	Low	P5 P6 P7 P8	4
11	cases	High	P4	1	41	autonomy	Low	P12 P13 P14	3
12	disaster	High	P8	1	42	exhaust	Low	P1 P2 P4	3
13	effect	High	P4	1	43	scrutiny	Low	P12 P13 P14	3
14	election	High	P14	1	44	smog	Low	P1 P3 P4	3
15	fees	High	P9	1	45	default	Low	P12	1
16	firm	High	P10	1	46	particles	Low	P1	1
17	progress	High	P5	1	47	radiation	Low	P7	1
18	scheme	High	P9	1	48	Brexit	unidentified	P9 P10 P11	3
19	tax	High	P11	1	49	Chernobyl	unidentified	P5 P6 P8	3
20	capabilities	Medium	P12 P14	2	50	haze	unidentified	P1 P2 P3	3
21	incentive	Medium	P9 P10	2	51	decontamination	unidentified	P5 P8	2
22	inhabitants	Medium	P5 P6	2	52	incense	unidentified	P1 P4	2
23	recovery	Medium	P5 P6	2	53	Arizona	unidentified	P11	1
24	senators	Medium	P12 P13	2	54	Arkansas	unidentified	P12	1
25	width	Medium	P2 P3	2	55	bloodstream	unidentified	P2	1
26	critics	Medium	P5	1	56	EU	unidentified	P9	1
27	diameter	Medium	P4	1	57	Newsroom	unidentified	P1	1
28	employee	Medium	P11	1	58	Pret a Manger	unidentified	P10	1
29	fragments	Medium	P1	1	59	subscriber	unidentified	P14	1
30	Index	Medium	P2	1	60	tsunami	unidentified	P8	1
	Number of lookups			46		Number of lookups			49
	Total number of lookups 46 + 49 = 95								

VERB				ADJECTIVE			
No.	Words	Classified by LC9000	Searched by	Frequency	No.	Words	Classified by LC9000	Searched by	Frequency
1	blamed	High	S1 S4	2	1	tiny	High	S3 S4	2
2	failed	High	S1 S2	2	2	available	High	S9	1
3	hired	High	S12 S13	2	3	effective	High	S11	1
4	reducing	High	S4 S10	2	4	financial	High	S10	1
5	celebrate	High	S4	1	5	harmful	Medium	S1 S2 S3	3
6	cost	High	S11	1	6	initial	Medium	S10 S11	2
7	forced	High	S2	1	7	vast	Medium	S5 S6	2
8	include	High	S9	1	8	unhealthy	Medium	S3	1
9	obtain	High	S10	1	9	wary	Low	S12	1
10	offer	High	S10	1	10	grainier	unidentified	S1 S2 S4	3
11	provide	High	S11	1	11	abandoned	unidentified	S5 S6	2
12	remains	High	S8	1	12	lunar	unidentified	S1	1
13	accused	Medium	S5 S6 S7 S8	4	13	toxic	unidentified	S3	1
14	fled	Medium	S6 S7 S8	3	14	quirky	unidentified	S12	1
15	issue	Medium	S12 S13 S14	3	Total number of lookup			22
16	flee	Medium	S5 S7	2					
17	posed	Medium	S13 S14	2					
18	restrict	Medium	S13 S14	2	ADVERB			
19	expand	Medium	S10	1	1	currently	High	S4	1
20	launch	Medium	S9	1	2	recently	High	S14	1
21	strengthening	Medium	S12	1	3	increasingly	Medium	S13	1
22	tackling	Medium	S1	1	4	partially	Low	S6	1
23	deemed	Low	S5	1	Total number of lookup			4
24	enroll	Low	S9	1					
25	hosing	unidentified	S1 S2 S3	3					
26	expediting	unidentified	S5 S7	2					
27	lip-synching	unidentified	S12	1					
Total number of lookup			44				

Corpus Linguistics Methods for Building ESP Word Lists, Glossaries and Dictionaries on the Example of a Marine Engineering Word List

Zorica Đurović, *Faculty of Maritime Studies, University of Montenegro, Kotor, Montenegro (zoricag@ucg.ac.me)*

Abstract: In addition to the general English knowledge required for nearly any human occupation today, vocabulary competence has been especially focused on seeking to keep pace with boosting Englishes for Specific Purposes. Owing to the possibilities offered by contemporary software solutions, corpus linguistics has been able to answer some specific questions on the vocabulary demand of texts, as well as to provide concentrated vocabulary lists according to their frequency in real-life texts (corpora). Aiming to provide our target learners of English for marine engineering purposes with a practical vocabulary tool to help them reach an adequate reading comprehension text coverage of 95%, we developed a marine engineering word list of 337 word families, accompanied by a list of 73 transparent compounds, which were derived from the corpus of marine engineering instruction books with 1,769,821 running words. The list can be studied in university classes or training courses for seafarers, through various types of vocabulary exercises, but it might also assist in building technical glossaries and dictionaries. The methodology used and procedures applied in the paper should hopefully be of assistance to other authors and language instructors working on other areas of technical English.

Keywords: VOCABULARY, WORD LIST, READING COMPREHENSION, CORPUS, INSTRUCTION BOOKS, ENGLISH FOR SPECIFIC PURPOSES, MARINE ENGINEERING, FREQUENCY

Opsomming: Korpuslinguistiese metodes vir die saamstel van ESD-woordelyste, glossariums en woordeboeke gebaseer op 'n skeepsingenieurswese-woordelys. Buiten 'n algemene kennis van Engels wat vandag vir byna enige mens se beroep vereis word, word daar vir 'n toereikende woordeskat daarop gefokus om tred te hou met die ondersteuning van Engels vir Spesieke Doeleindes. Weens die moontlikhede wat deur eietydse sagtewareoplossings gebied word, kon die korpuslinguistiek antwoorde op sommige spesifieke vrae oor die woordeskatvereistes van tekste verskaf en ook gekonsentreerde woordeskatlyste volgens frekwensie in werklike tekste (korpora) lewer. Met die strewe om aan ons doelleerders van Engels vir die vereistes van skeepsvaartingenieurswese 'n praktiese woordeskathulpmiddel te verskaf wat hulle kan help om 'n voldoende teksdekking van 95% vir leesbegrip te verkry, het ons 'n skeepsingenieurswese-woordelys van 337 woordfamilies, saam met 'n lys van 73 deursigtige samestellings, wat afgevoer is uit die korpus van handleidings vir skeepsingenieurswese met 1,769,821 lopende woorde, saamgestel. Hierdie lys kan in universiteitsklasse of opleidingskursusse vir seevaarders

deur middel van verskeie tipes woordeskatoefeninge bestudeer word, maar dit kan ook van hulp wees in die saamstel van tegniese glossariums en woordeboeke. Die metodologie wat gebruik word en die prosedures wat in hierdie artikel toegepas word, sal hopelik van waarde wees vir ander outeurs en taalonderwysers wat aan ander areas van tegniese Engels werk.

Sleutelwoorde: WOORDESKAT, WOORDELYS, LEESBEGRIP, KORPUS, HANDLEI-DINGS, ENGELS VIR SPESIFIEKE DOELEINDES, SKEEPSINGENIEURSWESE, FREKWENSIE

1.   Introduction

Although certain traces of language examination conducted upon collections of written texts can be found even from the age of antiquity, the notion of corpus linguistics in the modern sense has been related to the appearance of electronic corpora and other computing resources beginning in the 1970s (Nation 2016). Interestingly, the accelerated rise and application of this type of linguistic research has overlapped with the boosting of lexicography, all as a consequence of renewed interest in vocabulary and the new possibilities afforded by information technologies. The overall impact on language teachers has been two-fold. On the one hand, the more and more technical and demanding areas of English for Specific Purposes (ESP), accompanied by the massive development of information technologies, have imposed significant new challenges in terms of teacher competencies and course design. At the same time, new areas of research and methodologies have offered enormous opportunities in terms of the detailed computational analysis of abundant authentic material, as well as the testing and comparison of the obtained results, leading to more effective and learner-oriented language course material.

2.   Theoretical Background

The main ideas behind this renewed interest in vocabulary, perceived now as being in the "limelight of foreign language teaching and learning" (Gao and Liu 2020) is that "vocabulary is central to understanding and using a language at any level" (Hirsch and Coxhead 2009: 5), but that it also represents one of the major problems and most complex difficulty of any practical foreign language programme (Twaddel 1973: 61). Modern methodologies have provided researchers with the opportunity to "measure" vocabulary types and loads, as well as learning objectives. The central question to start with, therefore, regards the amount of vocabulary needed for adequate reading comprehension.

2.1   Reading comprehension

The quantitative vocabulary analysis of real-life texts (corpora) aims to answer

the question of the number of words needed to reach adequate reading comprehension, typically set at the levels of 95% and 98%. A "word" as a unit of measurement in this type of research most frequently stands for a word family, comprised of the headword with all its derivatives and inflected forms. For example, *read, reads, reading, reader, readable, readability*, and so on makes up one single word family, taking into account the learning burden, i.e. the amount of morphological knowledge the learners are expected to have (Nation 2016: 9).

The threshold of 98% was determined by Hu and Nation (2000) and also recommended by Carver (1994) and Kurnia (2003), claiming that 98% of known vocabulary (coverage of the text) or one out of 50 words unknown would be acceptable in order to understand a text adequately. Certain other authors, such as Laufer (1989), Laufer and Ravenhorst-Kalovski (2010) and Van Zeeland and Schmitt (2012) have advocated 95% as sufficient coverage of known words in a text in order to understand it correctly without additional assets and aids. Generally, authors have been guided by the recommendation summarized by Laufer and Ravenhorst-Kalovski (2010) by which an optimal reading comprehension would anticipate the ideal 98% lexical coverage of a text, which is usually achieved at a level of about 8,000 words, while the minimal adequate comprehension would be that of 95%, which we would expect to find at the level of 4,000–5,000 known word families, including proper nouns. The remaining 2–5% are expected to be guessed from the relevant context. Here, the type of vocabulary should be carefully considered, since, as has been shown by numerous studies involving ESPs, the desired thresholds in professional texts are barely reachable, or sometimes not reachable at all, even with all the available general and other available lists of English words; this, then, is the reason for the constant development of new specialized word lists (WL).

2.2   Word lists

Considering that both native and non-native speakers acquire vocabulary in the order of its frequency (Nation 2006), word lists have been used in language teaching for a long time, both as stand-alone lists and a part of textbooks and other teaching materials (Folse 2004; Nation 2006). General English (GE) lists provide the learners with the English words that are most frequently found in a variety of texts. Contemporary research into "specialized or technical vocabulary has focused primarily on producing word list of technical vocabulary in professional fields of expertise in English for Specific Purposes" (Coxhead and Demecheleer 2018: 84). Since the focus of our attention is a highly technical branch of ESP, we are following the general tendency and recommendation of upgrading the first 2,000 or 3,000 words with specialized vocabulary lists, which aim to reach the adequate reading comprehension threshold in the most efficient way.

For our specific research, we used the findings of previous research where

the target corpus was tested against the available general and relevant specialized word list for both its coverage ratio and in order to determine the lexical profile of the texts. These findings are explained in more detail in Section 4.1 Previous research findings.

2.3   English (vocabulary) for marine engineering purposes

Aiming to respond to the vocabulary needs of our target English language learners, that is, undergraduate students and trainees on Marine Engineering courses, we were led by their most practical needs in terms of language skills. Although English has become the *lingua franca* of almost all areas of international activity, this has been especially true in terms of maritime affairs, with English formally operating as its official language, as of the establishment of the International Maritime Organization in London in 1948. Consequently, in targeting our research objectives, we followed the expert advice and extensive teaching experience in the area. Even more importantly, we also followed the official requirements and recommendations made by the IMO's International Convention on the Standards of Training, Certification and Watchkeeping for Seafarers (STCW, Part 2.2) and the Model Course 3.17 — *Maritime English*, notably the part on Specialized Maritime English dedicated to marine engineering courses of English. Apart from general communication skills in terms of using internal communication systems, the majority of the language skills requirements (about 90% of the anticipated course and self-study hours) are dedicated to "Adequate knowledge of the English language to use engineering publications" (IMO Model Course 3.17 2015: 153). Led by these clear instructions, the area of our interest has been the reading comprehension of marine engineering publications, specifically instruction books.

3.   The Corpus

Instruction books have become an indispensable "tool" for the everyday onboard activities of marine engineers, being used in contexts ranging from familiarization with the ship's systems and machinery to regular maintenance activities and repairs. Therefore, an adequate reading comprehension of these publications is of utmost importance for the majority of their scope of activities, as well as for the entire shipping industry.

Following certain expert advice, primarily that of Chief Engineers, we sought to create a relevant selection of instructing engineering material for a tanker ship, a container ship, a cruise ship and an off-shore vessel. Additional material was added for the purpose of diversity and to cover up-to-date technologies, for example, in regard to propulsion, where we added instruction books for dual propulsion (both fuel oil and gas) and electric-driven engines. These additional types of instruction books and manuals were added in terms of the tanker ships deemed as the most numerous ones in worldwide fleets, but

also very similar to some others, e.g. container ships, especially as regards propulsion. Other material included technical manuals for various essential onboard machinery and systems, as presented in Table 1. In order to avoid any possibility of copyright or commerciality issues, we will not state the names of the manufacturers or the vessels. The final corpus material comprises thousands of pages of electronic material of items varying in length, converted to accommodate the software requirements.

Table 1: The corpus composition

Vessel type	Type of machinery	Number of files	Running words
TANKER SHIP	Main engine	8	466,237
	Operating media, auxiliary agents, machine elements	3	23,436
	Compressors	4	38,621
	Remote control valves	1	4,301
	Exhaust gas economizer	1	237,736
	Steering gear	1	32,330
	Incinerator	1	12,140
	Sterilizer	1	2,681
	Fresh water generator	1	2,128
	Air dryer	1	1,181
	Gas welder	1	3,645
	Calorifier	1	964
	Antifouling system	1	3,082
	Shaft earthing device	1	2,484
	Heat exchanger	1	3,402
	Control valves	1	5,012
	Local fire-fighting system	1	16,901
	Purifier	2	64,654
	Separator	4	14,645
	Bilge alarm monitor	1	6,977
	TOTAL	**36**	**942,557**
CONTAINER SHIP	Main engine	5	55,892
	Diesel generator	7	56,368
	Auxiliary boiler	1	156,370
	Centrifugal pumps	1	54,071
	Compressors	4	39,913
	Fresh water generator	2	25,725
	Separator	2	88,924
	22		**477,263**
CRUISE SHIP	Main engine	1	66,402
	Exhaust Gas Cleaning	2	26,679
	Waste water system	1	20,454
	Boilers	1	34,039
	Separation system	4	37,576
	Sanitation system	3	151,418
	TOTAL	**12**	**333,568**

SUPPLIER	Main engine and operating media	3	81,196
	Oil purifier	2	34,040
	Sea water desalinator	1	6,786
	Propulsor	1	5,313
	Safety instructions, environment	1	1,727
	Lubrication system	1	5,310
	Seal system	1	8,560
	Grease trap	1	6,622
	Generator sets	1	28,297
	Bilge water separator	1	8,271
	TOTAL	**13**	**522,690**
TOTAL		**83**	**2,279,078**

"The painstaking process" (Nation 2016: 224) of the additional cleaning of material (in relation to tables, references, brands, typos and the like) was applied to the best of the author's abilities, considering the huge amount of the material, originally found mostly in scanned formats. The prepared Corpus of Ship Instruction Books (CSIB) was finalized with 1,769,821 running words (tokens). Bearing in mind that we have a very technical and discipline-specific genre in question, we may say that our corpus is of representative size and content, so as to guarantee the validity of results and conclusions produced.

4.   Methodology

The initial method we used for our research is called Lexical Frequency Profiling (Laufer and Nation 1995). It provides authors and language teachers with the opportunity to analyse corpora in terms of vocabulary types and quantities and to test the coverage of available word lists, as well as to create new ones. The software that has recently been found most useful and convenient for this purpose is AntWordProfiler 1.4.0w, developed by Laurence Anthony (2014), as an upgraded version of the previously used RANGE programme (Nation and Heatley 1994).

As a starting point, we used the results of previous research on the same corpus (Đurović, Vuković Stamatović and Vukičević 2021) in terms of coverage by the existing genearal service and engineering word lists, aiming to prove the lexical demand of the corpus, as well as the need to build a specific marine engineering word list (MEWL). For the latter purpose, the same programme (AntWordProfiler 1.4.0w) was used for several reasons. First, it has proved itself to be one of the most appropriate programmes for building specialized vocabulary lists (or glossaries or dictionaries) since, in contrast to some other programmes used for building vocabulary lists, it can exclude the most frequent general vocabulary from further analysis, which is estimated to be known by LSP learners. Another reason is the value of having comparable results and being able to make reference to relevant findings and word lists built upon the same methodology. Finally, the Anthony's software is readily

available, quite simple to use for any language and offered free of charge.

In setting the cut-off point in the frequency count, we were guided by the goal of reaching the desired level of reading comprehension (of ship instruction books), which anticipated a minimum of 95% of known vocabulary. With a list which was estimated to be of a size to represent an attainable task for language learners, taking into account the available time for learning, the reached level of 95% corpus coverage (with the most frequent GE words and the MEWL) would at the same time be its positive evaluation result (Dang and Webb 2016: 133).

For formatting the lists into headwords only, lemmas or expanding them into all-family-members form, we used the Familizer + Lemmatizer programme (Cobb 2018). For corpus preparation and converting it into "plain text" format, as required by the software, we used AntFileConverter (Anthony 2017).

Additionally, for detecting the concordances of a certain word within the corpus, the AntConc software (Anthony 2019) can be used. Owing to this kind of software programmes, each word from our list can be checked for concordances and n-grams to be included in a glossary or dictionary.

4.1   Previous research findings

In examining the target corpus, we first analysed its vocabulary load and types in terms of reaching the desired reading comprehension. Aiming to provide precise answers as to the quantity and types of vocabulary needed for the purpose and following the methodologies used by recognized authors in the field, these questions were the focus of our previous research (Đurović, Vuković Stamatović and Vukičević 2021). As such, we here briefly present the lexical profiling methodology and the results, since they clearly point to the need for a specialized marine engineering word list and provide solid justification for our current research.

In particular, in testing our corpus against the General Service List (West 1953) and the Academic Word List (Coxhead 2000), which are usually applied together in this kind of research, the cumulative coverage amounted to 79.46% (71.39% + 8.07%, respectively), which is lower than the average of 86.1% found for academic texts (Nation 2000: 27). Considering general English only, the coverage level of 71.39% is significantly below the usual coverage of 78–98%, as reported for various types of written texts (e.g. Nation and Waring 1997). This generally means that, with knowing only 2,000 first English words, even with the most common academic words, about every fifth word of the text remains unknown (20.54%), which would make both reading and understanding the text very difficult.

When calculating the total amount of general English words needed for adequate comprehension, using the more contemporary Nation's word lists extracted from the huge corpora of the British National Corpus and the Corpus of Contemporary American (BNC/COCA, available at https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists), a coverage of 95%

was reached not sooner than with the 12,000 most frequent English words, whilst 98% coverage is not reachable even with all the available 25,000 English words, including the additional four lists of proper nouns, abbreviations, transparent compounds and marginal words. This means that a clear understanding of this kind of text is almost unattainable, even for a native speaker, without an engineering background.

In taking into consideration the available word list from adjacent fields and genres, the only two available lists which we found to be relevant and appropriate to this kind of research were Ward's Basic Engineering English Word List (BEEWL) (Ward 2009), and Hsu's Engineering English Word List (EEWL) (Hsu 2014). Both lists were extracted from undergraduate textbooks appropriate to various engineering fields. In addition, both of them showed lower coverage in our target corpus (13.53% compared to 16.4%, and 10.11% compared to 14.3%, respectively). These results are also understandable considering the very technical nature of our target corpus, when compared to that of textbooks, which are more narrative in nature.

Bearing all these findings in mind, we sought to create a specialized marine engineering word list, aiming to reach the adequate reading comprehension level for marine engineering instruction books. Taking into consideration that this kind of genre is abundant in both diagrams and tables, as well as supplemental explanations and abbreviations accompanying these materials, we opted to place the adequate reading comprehension threshold at 95%.

5.   The Marine Engineering Word List

Carefully examining the best way to apply the methodology stated above in our research so as to be relevant, efficient and justified, we took several points into account. Firstly, our target learners are either university-level students or active seafarers undertaking professional training. Furthermore, bearing in mind the fact that English is learned in many countries from an early age, we deemed that it would be reasonable to expect them to be competent in reading and understanding at least 3,000 basic English words (BNC/COCA), especially considering that it refers to the receptive knowledge and not necessarily productive language skills. This follows certain recent trends in expending the high-frequency list from 2,000 (West 1953; Nation 2001) to the first 3,000 English words (Schmitt and Schmitt 2014; Nation 2016). Additionally, adequate English proficiency is mostly required and tested by shipping companies during their recruitment procedures.

Finally, since in many countries, officers' training courses do not need to be organized through university courses, but can be organised through any certified training centre, and taking into account the fact that this type of text is more technical than academic (although the AWL list coverage of 8.07% is not to be overlooked), we decided to head directly to the marine engineering word

list (MEWL) for obtaining early specialization (Coxhead and Hirsch 2007) in English for Marine Engineering Purposes (EMEP).

6.   Data analysis and results

In determining the frequency threshold for our list, we took into account the final objective of 95% coverage, the relevant lists to test against (and those to exclude from the count) and the reasonable list size. In the end, we opted for a frequency of at least 50, so the final list was formed of 337 headwords (see Addendum 1) with 73 transparent compounds (see Addendum 2). We did not apply the range and dispersion criteria here, since they are more relevant to huge corpora and extracting common vocabulary for various professional areas (e.g. Coxhead and Hirsch 2007) than one specific professional profile (e.g. Coxhead and Demecheleer 2018). In our research, the corpus comprises a variety of technical materials in terms of different machinery and ship's systems, all nearly equally important to the single occupation of marine engineers.

Aiming to obtain the most precise results possible, in analysing the corpus, we extracted the most frequent abbreviations and marginal words (the latter mostly typos and conversion errors), and added them to Nation's original lists, respectively. This selection process is never easy or perfect, since, for example, some abbreviations are recognizable (e.g. *cyl* for *cylinder*, *hfo* for *heavy fuel oil*), but sometimes it is difficult to distinguish between an abbreviation and a typo or conversion error. Fortunately, they were not numerous, owing to the carefully conducted cleaning process of the corpus, and, once added to one of the additional Nation's list, cumulatively, they do not affect the final results, apart from enabling the production of a "purer" word list.

In addition, in spite of our efforts to initially remove as many proper names as possible, so as to avoid the commerciality of the data, what remained was also added to the Nation's list of proper nouns. Taking into account that all the words considered above are expected to be easily understood and recognized from the context, thus not bearing a learning burden of significance, we excluded them from our further frequency analysis together with the most frequent GE words.

After having a close look at the Nation's list compounds, we deemed it too abundant for our target learners and significantly more difficult to learn compared to the most frequent general English words, thus we did not exclude it from our initial count. On the contrary, we decided to make a separate list of transparent compounds (Addendum 2) derived from the corpus itself. In that, we followed the recommendation of having a separate list of this type of vocabulary (Nation 2016: 70). Some of the items overlap with those from the Nation's list (e.g. *setpoint, standstill*) and the rest originate from the corpus (e.g. *sootblower, crankthrow*). Hyphenated forms are not frequent in this type of text, and they were initially eliminated by dividing their constituents into sepa-

rate words. In addition, those compounds that could not be easily understood according to their constituents, were left as part of the initial list, since they cannot be deemed transparent (e.g. *bulkhead*). Here we must note that, although presented separately, they should be taken as one with the initial list of words, since they are of equal significance to both vocabulary skills and course designs and materials.

The final results are presented in Table 2.

Table 2:   Coverage of MEWL in the CMEIB

Word lists	Tokens	Coverage (%)
BNC/COCA 3,000 + proper nouns, abbreviations and marginal words	1,547,067	87.41%
Transparent compounds	12,806	0.72%
MEWL without compounds	130,994	7.41%
—	78,954	4.46%
Total	1,769,821	100%

As we can see, the coverage of the first 3,000 most frequent English words (with the additional three lists, as usually excluded first in the analysis) amounts to 87.41%. If we compare this to other findings, we can see that the general English coverage is lower in our target corpus than, for example, in business research articles, where 3,000 words with proper nouns cover as much as 90.84% of the material, in business textbooks, where the coverage is 94.15% (Hsu 2014: 251), in popular science books where 3,000 words with the additional 4 Nation's lists cover 92.65% (Vuković Stamatović 2020) or in various types of texts where 3,000 word coverage is of the order of 90–93% (Nation 2006). The coverage is lower even compared to the coverage of the first 2,000-word BNC/COCA lists in some types of texts, which ranges from about 79% (in various academic text) to 89% in school journals and novels (Nation 2006, 2016; Fraser 2007). This clearly points to the complexity of our target vocabulary in terms of its technicality. It also speaks loudly in favour of the need for a specialized word list to accommodate adequate coverage of the target genre.

For the purpose of illustration, we can use another option of the programme to see the corpus text coloured as per the word lists (Figure 1), where (in the extract presented), for example, pink stands for the first 3,000 BNC/COCA words, orange for MEWL, blue for the abbreviations, while the words outside the list are left black.

1.2.1 Oil and atomizing systems
The atomizing medium and oil are supplied from valves through flexible hoses to the burner lance.
The automatic shut-off valves, re-circulation valve, solenoid valves, and the rest of the
necessary burner mountings are all fixed on the wind-box. All shut-off valves are
standard ball valves, and the automatic valves are with electric/pneumatic actuators.
The arrangement of the components on the burner is made as simple as possible and
is therefore very user friendly regarding operation and maintenance.
The oil flow control valve is fitted together with a pneumatic converter/positioner. To
obtain a very accurate regulation of the oil flow, which is very important to have a large turn-down ratio, the valve is also fitted with an adjustable Cv-adjuster. This Cv-adjuster makes it possible to adjust the pressure loss coefficient of the valve whereby the oil flow regulation can be optimised to each single installation. The oil flow is measured by a screw-type flow meter.

Fig. 1:   Level lists presented in colours

We can also present MEWL with compounds only using the same option, where green marks the MEWL words and transparent compounds are given in red, while the remaining are left black.

Proximity sensors per cylinder, i.e. one for each guideplane are placed on brackets typically mounted on
the Exhaust side structure below both guide shoes.

The system monitors the variation in distance between sensor and guide shoe bottom surface, in bottom dead centre BDC. The measure- ments represent the relative wear in all three crank-train bearings, simultaneously.
. Seawater Cooling System . . . .

Fig. 2:   MEWL and compounds' list presented in colours

This option could also be used for text glossing, which is often helpful in terms of reading technically demanding texts such as ours (Nation 2013). In this way, in addition to the glossaries frequently added to the textbooks, the learning materials can be compiled out of authentic texts (whether adjusted or not) with

detected and glossed technical terms, followed by a definition, translation, or similar.

The coverage of the "pure" marine engineering word list of 337 word families is 7.41%, i.e. 8.13% adding the list of 73 transparent compounds. The purity of the final list is attained by additional analysis of the list and detecting additional members to add to the Nation's original 3,000 families (e.g. *purifier, cleanable, abnormal*). Moreover, during the conversions, some technical words are not recognized, thus additional attention should be paid to "unclassified" words, such as e.g. *bunker (bunkering)*, *alignment (misalignment)*, and similar terms, which were subsequently added to the list.

In total, together with the first 3,000 GE words, proper nouns, abbreviations and marginal words, a level of 95.54% was reached, thus fulfilling our goal of attaining the adequate reading comprehension threshold. Taking into consideration that the desired level could be attained with not less than 12,000 general English words only, our final results perfectly fit the findings of Laufer and Ravenhorst-Kalovski (2010), by which the threshold of 95% is expected to be reached through the use of 4,000–5,000 word families.

The remaining share of the unknown vocabulary in our corpus would here form below 5% (4.46%), which means that the adequate reading comprehension threshold is comfortably reached. Adding the introductions, professional knowledge and abundance of schema and diagrams, the understanding of this kind of professional text should therefore no longer present any significant difficulties. However, an aid such as a glossary or dictionary for instruction books would certainly be welcomed by both (future) marine engineers and teachers of this challenging branch of ESP.

7.   Glossaries, dictionaries and pedagogical implications

Having a ready-made list of the most frequent vocabulary extracted from marine engineering instruction books, the course material designers can then choose whether to use them as additional material, compile new textbooks and exercises with excerpts and examples taken from authentic instruction books, or else use the list in any other way they find suitable. Given the size of the list, it might form a convenient and achievable task for (a part of) university studies, where about 800 new words are considered a practical learning goal for each two years of study (Dang and Webb 2106: 174). Moreover, additional corpus linguistics methods can be introduced to students as well, and thus they could be required to find concordances, n-grams or full examples for the word use from the corpus itself.

In addition, the list would be a more than useful tool for English courses held at seafarer training centres, since the attendees would generally have some experience of using onboard instruction books. Furthermore, any marine engineer would be able to make good use of having such a concentrated list of specialized vocabulary at hand during their onboard service. The list, therefore,

could help in creating a monolingual or bilingual glossary, with or without pronunciation transcription.

Maritime English textbooks, like those of many other areas of ESPs, are often supplemented with a glossary, frequently a bilingual one (e.g. Spinčić and Pritchard 2009: 180-242) which covers the textbook syllabus. For this purpose, it would be sensible to use authentic texts, such as excerpts from various instruction books, and the word list might be divided into sub-lists or learning chunks in order to suit the organization of classes and exams. Frequently, textbooks cover more than one semester of studies or are printed in several volumes, which can be covered by a common glossary. Additionally, the list might also be added to existing glossaries to cover combined class material.

Considering stand-alone glossaries, which are frequently found online as useful links provided by various maritime organizations, institutions and companies, we first noticed the scarcity of specific marine engineering glossaries, in contrast to nautical or (general) maritime ones. The most of maritime glossaries are generally very simple in form. They most frequently offer just the target term or phrase with a simple context explanation (monolingual) or translation (bilingual), occasionally including additional details such as the phonetic transcription (sometimes offering both British and American variants) or phrases and collocations.

The extraction of the most frequent words is also applicable to a more comprehensive and demanding project, such as a marine engineering dictionary, whether that be creating a new one or adding to an existing one to make it more contemporary and comprehensive. Almost any dictionary has been built upon a corpus (Nation 2016: 176), and using the possibility of measuring the (lemmatized) vocabulary frequency range within it would certainly be an advantage in this regard (De Schryver and Nabirye 2018). Compared to word lists, which are intended to be learned for the purpose of unaided reading, the cut-off point for dictionaries can be moved to lower frequencies, so as to be more inclusive in terms of technical (low-frequency) vocabulary. A further advantage of the proposed methodology for building a technical glossary or dictionary is that, upon compiling a corpus of adequate composition and size, the building of the macrostructure would be solidly founded upon the frequency of the vocabulary beyond the most frequent GE words, the lists of which can be excluded from further analysis.

This kind of dictionary might also be supplemented by illustrations (e.g. Carić 2011) and/or examples from the corpus. This might prove especially useful, since using dictionaries to explore aspects of target words has been shown to be one of the most efficient ways of learning about vocabulary within the language awareness strand of learning strategies (Hirsch and Coxhead 2009; Macalister and Nation 2011). Here we must note that a technical glossary or dictionary would mostly not contain general English meanings and examples. For example, the word *pinion* would be here referred to as a type of gearwheel. However, when deemed necessary to include both, the notation GE (general

English) or ME (marine engineering) would follow (e.g. *average* (GE) and *average* (ME)). More considerations on polysemous words in marine engineering are given in Section 8. Limitations of the study.

The lists of most common abbreviations should be added to the dictionary, which might also include more relevant details, such as the marked part of speech and further examples, or perhaps illustrations. The words can be followed by the frequency mark, such as, for example, in Macmillan's Dictionary (2012), bearing in mind the specific corpus it was derived from.

For the purpose of generality and following the organization of various dictionaries, we offer a monolingual simplified example for the most frequent word (appearing 10,871 times) in the instruction books:

(1)　**valve** /vælv/ n. – a device that opens and closes to control the flow of a liquid or gas

Following the main term and definition, AntConc (Anthony 2019) (or another programme offering a similar function) provides an excellent opportunity to check the concordances of certain terms from the corpus. Noun phrases and certain collocations containing the target term usually follow the head definition.

(2)　**valve** /vælv/ n. –
　　　by-pass valve –
　　　check valve –
　　　flow control valve –
　　　motor-oil flow valve –
　　　pressure valve –
　　　return valve –
　　　safety valve –
　　　steam valve –
　　　stop valve –
　　　oil sample valve – , and so on.

To avoid headword repetition in multi-word units, it is sometimes replaced with a sign such as ~ (e.g. Rapovac 2002; Spinčić and Pritchard 2009).

(3)　**valve** /vælv/ n. –
　　　by-pass ~ –
　　　check ~ –
　　　flow control ~ –, and so on.

With noun phrases where the term in question is not the head noun, these examples often come first, preceding those as in Example (2).

(4)　**valv**e /vælv/ n. –
　　　valve block –
　　　valve body –

> valve bottom –
> valve bypass –
> valve components –
> valve control –

Again, these can be written as:

(5)   **valve** /vælv/ n. –
        ~ block –
        ~ body –
        ~ bottom –, and so on.

In addition to collocations, dictionaries often contain authentic examples of the use of the words within phrases and sentences, which can easily be found using the same programme. For example:

(6)   by-pass valve – ex. Adjust the oil pressure to a suitable level on the by-pass valve.

Comparing the means of presentation, word families are generally used for word lists for practical reasons, whilst in practice they are expanded, usually to include all family members when used for the programme analyses presented above. For dictionaries (and glossaries) it is more convenient and useful to use lemmas, so that different parts of speech can be presented as separate items (for example *actuate v.*, *actuation n.*, *actuator n.*). For this purpose, again Familizer + Lemmatizer (Cobb 2018) might be used. For example, the single word family *bear* comprises 12 family members (word forms), whilst, for instance, in a bilingual maritime dictionary (Rapovac 2002) we can find the meanings and examples of *bear*, *bearer* and *bearing*, in various collocations and contexts. Based on the proposed software being used, we should note here that the selection should rely upon the frequency of tokens converted directly into lemmas (and accompanied by compounds and phrases) in order to reduce the interference of the software in the selection process during the list expansion and reduction processes.

Further details on the organization of dictionaries vary according to their intended purposes and the choices made by the authors. Compared to the painstaking endeavour of finding relevant examples and the collection and processing of abundant material manually, such as, for example, was the case with the quotation slips used for the first Oxford English Dictionary, the idea of this paper was to present some of the available IT-supported methodologies of corpus linguistics, which make the process incomparably easier and faster, but also much more accurate, thus providing legitimacy to the use of this system.

8.   Limitations of the study

Despite the best efforts of the author to apply the most accurate methodology

and make the best possible and reasonable decisions along the way, we accept that neither the methodology nor the results are ever perfect (Nation 2016: 182). For example, the conversion of large instruction books, some running to as many as over 800 pages of scanned material, resulted in many typos and conversion errors. This added to the words outside the lists, whilst a number of those should add to the frequencies within them.

The programme itself does not recognize different spelling options (e.g. *authorized* and *authorised*), and thus counts them as separate words with separate frequencies, where the frequency should in fact be cumulative. Moreover, multiword units are always a point of issue in this kind of research (Nation 2016). For example, some of them can be written separately (e.g. *cam shaft*), and some as a single word (*camshaft*), which makes the statistics in terms of frequency not entirely precise. Additionally, careful attention should be paid to the words "unclassified" by the programmes, which are often the "missing" members of the general English word families or unrecognized technical words. In particular, general programmes such as Familizer + Lemmatizer do not recognize certain words from the marine engineering register, such as, e.g. *crosshead* or *bedplate*, and thus we had to retrieve many of them from the "unclassified" word types and manually add them to the lists used or produced by the programme.

When it comes to such a technical branch of ESP as marine engineering English is, additional attention should be paid to semantic issues. In particular, a certain portion of the words could be classified as belonging to Step 3 of a four-step rating scale for technical vocabulary (Chung and Nation 2003: 104), being referred to as polysemous or cryptotechnical words (Fraser 2009). This means that some of the words classified as belonging to the most frequent general English words often have a completely new and technical meaning in marine engineering contexts, either individually (e.g. *wear, draught, average*) or in collocations (e.g. *jacket water cooling system, guide shoe*). In the subsequent table, we present some examples from the first 3,000 GE words which have different meanings in marine engineering English, either individually, or in collocations. The definitions are somewhat shortened and simplified for practical reasons of presentation.

Table 3:   Polysemous words from the first 3,000 GE (BNC/COCA) words

The first 1,000 (*BNC/COCA*, all family members)	General English	English for Marine Engineering Purposes
actuate v. *actuator n.*	(added to the word family *actual*)	start, move mover, starter
rocker n. *arm n.*	someone who plays or likes rock music one of the two long parts of our body ending with a hand	*rocker arm* — a rocking lever of the valve opening mechanism

bearing n.	the particular way in which someone stands, moves, or behaves	a machine element that constrains relative motion to only the desired motion
excite v.	make someone feel happy and enthusiastic	to start an electric motor
film n.	movie	a thin (oil) layer
ground v.	earth; touch the ground	touch seabed with the bottom of the ship
mean adj.	evil	average value
mess n. *room n.*	disorder space	*mess room* — ship's dining room
running n.	moving quickly; controlling a business or organization	engine operation
shoe n.	something you wear on your feet	guide shoe — a piece of a crosshead guide
skirt n.	a piece of female clothing	cylindrical walls of a piston
The second 1,000 (*BNC/COCA***, all family members)**	**General English**	**English for Marine Engineering Purposes**
arrestor, arrester n.	(added to the family *arrest*)	a device preventing the unwanted transmission of electricity
average n.	the mean	damage to the ship
breather n.	(added to the family *breathe*)	ventilation opening or (relief) valve
guide n.	an informative book or person guiding you	a device that directs the motion of a machinery part
nut n.	a type of food	a round hollow metal piece used with screws
pin n.	a small thin piece of metal with a sharp end	*gudgeon (piston) pin* — a round metal piece connecting the piston to the piston rod *crank pin* — a round metal piece for crankshaft connection
stress n.	a worried or nervous feeling	phisical pressure on a material or vessel (hull)
trunk n.	the main part of a tree the braches grow out of	trunk (piston engine) — an engine without a crosshead
jacket n.	a piece of clothing	a protective or insulating cover around the cylinder
The third 1,000 (*BNC/COCA***, all family members)**	**General English**	**English for Marine Engineering Purposes**
draught n.	cold air that blows into a room	the distance from the water line to the bottom of the vessel
survey n.	a set of questions designed to explore public opinion	a technical inspection
surveyor n.	a person asking a set of questions to explore public opinion	a person conducting technical inspections and examinations

Here again, we have to mention the conditionality of classification, since many of the technical terms are shared among various fields, in this case, especially with general and other branches of engineering. Overlaps are not a rarity in the produced word lists, which was the reason that some authors have decided to build common core vocabulary word lists, such as the Academic Word List (Coxhead 2000), the Science-Specific Word List (Coxhead and Hirsch 2007), and other similar lists.

When it comes to bilingual formats, things get even more complex, depending on the lexicographical features of the local language in relation to the maritime sector, which, most usually anticipates the colloquial use of numerous Anglicisms and words borrowed from Italian or Spanish, for example. For all the reasons mentioned above, this kind of analysis cannot be left to pure statistical processing; it is evident that human (expert) intervention of close account and expertise is in fact required every step of the way.

Considering the corpus itself, for university students, for example, it can be broadened by the use of textbooks and possibly scientific articles, although the latter can make a separate academic corpus to examine in terms of lexis. The reason for this could be the contents of the list we produced. We can easily spot words from outside the marine engineering field, but which are related to, for example, physics (e.g. *amplify*), medicine (e.g. *diaphragm*), electrical engineering (e.g. *coil*), and so on. Therefore, the list as is could be named more precisely as "the word list of ship instruction books", and new and additional marine engineering word lists might be developed from a wider or different corpus.

In order to overcome the possible shortcomings of the overall process, we invested out best efforts in following and taking into consideration various previous findings and recommendations. Among those, we tried to explain the application of the methodology and the decisions made along the way. This provides justification and makes the results clearer, but might also help other authors in building relevant specialized lists of other technical areas and, indeed, other ESP learners.

9.   Conclusion

Analysing the most practical language needs of our target language learners — in this case future and active marine engineers following English for Marine Engineering Purposes courses, we embarked on the ambitious project of collecting, selecting and analysing their key corpus of marine engineering instruction books in terms of both vocabulary types and the overall lexical burden. Following certain previous research findings, we applied the lexical profiling methodology and some of the most updated software for the creation of a specialized marine engineering word list. The final list comprises 337 word families plus 73 compounds and cumulatively covers 8.13% of the target corpus. Together with the 3,000 most frequent English words (BNC/COCA) and an

additional and broadened lists of proper nouns, marginal words and abbrevia-tions, we succeeded in reaching the desired coverage level of 95%, more pre-cisely, 95.54%. This means that with a knowledge of the 3,000 most frequent English words and familiarization with a range of the most frequent proper names and abbreviations, with MEWL, less than 5 out of 100 words of this demanding type of technical material should remain unfamiliar to the reader, a percentage which should not significantly affect the comprehension of the text. Compared to the about 12,000 general English words initially needed to achieve a similar result, the amount of vocabulary required is incomparably reduced, and as such, the list we have produced should certainly provide for early specialization in English for Marine Engineering Purposes.

We have also put additional efforts into explaining the methodology and its application, and providing explanations and justifications for the decisions made during the (replicable) process and the final cut-off points, in order to make our own modest contribution to future research in the field and the enhancement of the available methodology. Moreover, we sought to provide practical suggestions for building (marine engineering) glossaries and diction-aries based upon a similar methodology, which will, hopefully, be a matter of interest for other ESP teachers and lexicographers, just as it is for the author of this work.

References

Anthony, L. 2014. *AntWordProfiler* (Version 1.4.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available at: https://www.laurenceanthony.net/software.

Anthony, L. 2017. *AntFileConverter* (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available at: https://www.laurenceanthony.net/software.

Anthony, L. 2019. *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available at: https://www.laurenceanthony.net/software.

Carić, T. 2011. *Englesko crnogorski pomorski slikovni rječnik*. Tivat: Fakultet za mediteranske poslovne studije.

Carver, R.P. 1994. Percentage of Unknown Vocabulary Words in Text as a Function of the Relative Difficulty of the Text: Implications for Instruction. *Journal of Reading Behavior* 26(4): 413-437.

Chung, T.M. and I.S.P. Nation. 2003. Technical Vocabulary in Specialised Texts. *Reading in a Foreign Language* 15(2): 103-116.

Cobb, T. 2018. From Corpus to CALL: The Use of Technology in Teaching and Learning Formulaic Language. Siyanova-Chanturia, A. and A. Pellicer-Sanchez (Eds.). 2018. *Understanding Formulaic Language: A Second Language Acquisition Perspective:* 192-211. New York: Taylor & Francis.

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.

Coxhead, A. and M. Demecheleer. 2018. Investigating the Technical Vocabulary of Plumbing. *English for Specific Purposes* 51: 84-97.

Coxhead, A. and D. Hirsch. 2007. A Pilot Science-Specific Word List. *Revue Française de Linguistique Appliqueé* 12(2): 65-78.

Dang, T.N.Y. and S. Webb. 2016. Making an Essential Word List for Beginners. Nation, I.S.P. (Ed.). 2016. M*aking and Using Word Lists for Language Learning and Testing*: 153-167. Amsterdam: John Benjamins.

De Schryver, G.-M. and M. Nabirye. 2018. Corpus-driven Bantu Lexicography. Part 2: Lemmatisation and Rulers for Lusoga. *Lexikos* 28: 79-111.

Đurović, Z., Vuković Stamatović, M. and M. Vukičević. 2021. How Much and What Kind of Vocabulary do Marine Engineers Need for Adequate Comprehension of Ship Instruction Books and Manuals? *Círculo de lingüística aplicada a la comunicación* 88 (in print).

Folse, K. 2004. *Vocabulary Myths: Applying Second Language Research to Classroom Teaching.* Ann Arbor, MI: University of Michigan Press ELT.

Fraser, S. 2007. *Providing ESP Learners with the Vocabulary they Need: Corpora and the Creation of Specialized Word Lists.* Retrieved from: https://ir.lib.hiroshima-u.ac.jp/files/public/2/24812/20141016150057516129/h-gaikokugokenkyu_10_127.pdf.

Fraser, S. 2009. Breaking Down the Divisions Between General, Academic, and Technical Vocabulary: The Establishment of a Single, Discipline-based Word List for ESP Learners. *Hiroshima Studies in Language and Language Education* 12: 151-167.

Gao, J. and H. Liu. 2020. Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners. *Lexikos* 30: 111-142. http://lexikos.journals.ac.za; https://doi.org/10.5788/30-1-1548.

Hirsch, D. and A. Coxhead. 2009. Ten Ways of Focussing on Science-specific Vocabulary in EAP Classrooms [online]. *English Australia Journal* 25(1): 5-16. Available at: https://search.informit.org/doi/10.3316/informit.473250020621629.

Hsu, W. 2014. Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates. *English for Specific Purposes* 33: 54-65.

Hu, M. and I.S.P. Nation. 2000. Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language* 13(1): 403-430.

Kurnia, N. 2003. *Retention of Multi-word Strings and Meaning Derivation from L2 Reading.* Unpublished Doctoral Dissertation. Wellington: Victoria University of Wellington, New Zealand.

Laufer, B. 1989. What Percentage of Text-lexis is Essential for Comprehension? Lauren, C. and M. Nordman (Eds.). 1989. *Special Language: From Humans Thinking to Thinking Machines:* 316-323. Clevedon: Multilingual Matters.

Laufer, B. and I.S.P. Nation. 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16(3): 307-322.

Laufer, B. and G.C. Ravenhorst-Kalovski. 2010. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language* 22(1): 15-30.

Macalister, J. and I.S.P. Nation. 2011. *Case Studies in Language Curriculum Design: Concepts and Approaches in Action Around the World.* New York: Routledge.

Maritime English. 2015. Model Course 3.17. London: International Maritime Organization.

Nation, I.S.P. 2000. Review of *What's in a Word? Vocabulary Development in Multilingual Classrooms* by N. McWilliam. 2012. *Studies in Second Language Acquisition* 22(1): 126-127.

Nation, I.S.P. 2001*. Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I.S.P. 2006. How Large a Vocabulary is Needed for Reading and Listening? *Canadian Modern Language Review* 63(1): 59-82.

Nation, I.S.P. 2012. *The BNC/COCA Word Family Lists.* Available at: http://www.victoria.ac.nz/ lals/about/staff/paul-nation (accessed January 2019).

Nation, I.S.P. 2013. *Learning Vocabulary in Another Language.* Second edition. Cambridge: Cambridge University Press.

Nation, I.S.P. 2016. *Making and Using Word Lists for Language Learning and Testing.* Amsterdam: John Benjamins.

Nation, I.S.P. and A. Heatley. 1994. *Range: A Program for the Analysis of Vocabulary in Texts (Software).* Retrieved from: http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx (accessed January 2019).

Nation, I.S.P. and R. Waring. 1997. Vocabulary Size, Text Coverage, and Word Lists. Schmitt, N. and M. McCarthy (Eds.). 1997. *Vocabulary: Description, Acquisition and Pedagogy*: 6-19. Cambridge: Cambridge University Press.

Rapovac, D. 2002. *Pomorski rječnik, englesko–srpski.* Kotor: Fakultet za pomorstvo.

Rundell, M. 2012. *Macmillan English Dictionary for Advanced Learners*. 2012. Second edition. Oxford: Macmillan Education.

Schmitt N. and D. Schmitt. 2014. A Reassessment of Frequency and Vocabulary Size in L2 Vocabulary Teaching. *Language Teaching* 47(4): 484-503. doi:10.1017/S0261444812000018.

Spinčić, A. and B. Pritchard. 2009. *An English Textbook for Marine Engineers II.* 4th edition. Rijeka: Pomorski fakultet.

Twaddell, F. 1973. Vocabulary Expansion in the TESOL Classroom. *TESOL Quarterly* 7(1): 61-78. doi:10.2307/3585510.

Van Zeeland, H. and N. Schmitt. 2013. Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics* 34(4): 457-479.

Vuković Stamatović, M. 2020. Vocabulary Suitability of Popular Science Books for English for Science Classes: A Corpus-Based Study. Čubrović, Biljana (Ed.). 2020. *BELLS90 Proceedings: Belgrade English Language and Literature Studies. International Conference to Mark the 90th Anniversary of the English Department, Faculty of Philology, University of Belgrade. Volume 1:* 439-447. Belgrade: Faculty of Philology, University of Belgrade.

Ward, J. 2009. A Basic Engineering English Word List for Less Proficient Foundation Engineering Undergraduates. *English for Specific Purposes* 28(3): 170-182.

West, M. 1953. *A General Service List of English Words*. London: Longman, Green & Co.

Addendum 1: The headwords of the marine engineering word list (without transparent compounds)

aboard	bracket	crank	erosion
abrasive	brass	cylinder	evaporate
accessory	bronze	debris	ex
accord	buffer	decant	expire
acid	bulkhead	default	fasten
actuate	bunker	defect	fax
acute	cabin	deflect	fecal
adhesive	calibrate	deform	fixture
adjacent	calorific	detergent	flange
adsorb	cam	deteriorate	flap
aft	cartridge	deviate	flotation
align	casing	diagnosis	flue
alkaline	caterpillar	diagram	fluid
alternate	caution	dial	flush
aluminium	cavity	diameter	fore
ambience	centrifuge	diaphragm	foul
amplify	centripetal	diesel	friction
analog	certify	differential	furnace
annex	chlorine	digit	fuse
annular	clamp	din	gasket
anode	classification	dip	gauge
anti	clog	discard	generator
appendix	clutch	discrete	geometry
appliance	cock	disinfect	gland
arrow	coil	dismantle	globe
ash	collar	dispense	glove
assemble	combustion	displace	grease
astern	communicable	dissolve	grind
automate	compartment	distillate	groove
automobile	compatible	drip	gudgeon
auxiliary	comply	droop	halogen
axis	compress	dual	hammer
barge	con	duct	harness
barrel	condense	durable	hexagon
batch	cone	duration	hoist
bellow	configure	dynamic	hood
bilge	console	effluent	horizontal
blade	contaminate	eject	hose
blink	copper	electrode	hub
bolt	corrode	emulsion	hull
bonnet	crane	enclose	humid

hydraulic	nominal	rinse	sulphur
icon	nozzle	rod	sump
identical	offset	rotate	surge
idle	optimum	rubber	surveillance
ignite	opus	rudder	synthetic
impel	orifice	rust	tab
impulse	oxidation	saline	tag
incinerate	oxygen	sanitize	telescope
inert	parameter	satisfactory	terminal
inhibit	particle	sauer	thermal
insulate	paste	scavenge	thermostat
intact	permissible	scrape	thread
integral	pinion	scrubber	threshold
intercept	piston	seizure	throttle
interface	pliers	sensor	thrust
interval	plunge	serial	tiller
jacked	pneumatic	servo	tilt
kit	polyamide	shaft	tolerance
knob	polymer	silicon	torch
lance	potable	sketch	torque
lever	precaution	sleeve	torsion
linear	preface	sludge	toxic
linen	preliminary	socket	transducer
liner	prescribe	sodium	turbine
litre	preset	soiled	tween
loop	prolong	solenoid	usage
lube	propel	solvent	utensil
lubricate	propulsion	soot	vacuum
manifold	proximity	spa	valve
manoeuvre	pulley	span	vane
manometer	pulse	spanner	velocity
membrane	puncture	spark	ventilate
mesh	purge	specimen	verify
micro	quarantine	spindle	vertical
mineral	rack	spiral	vibrate
minimise	radial	splash	viscous
mist	ram	spool	volt
moisture	ramp	stack	warranty
molybdenum	recreation	static	weld
mount	refract	steer	wrench
multi	relay	stool	yoke
needle	residue	stud	
nipple	resilience	stuffed	
node	rim	suction	

Addendum 2:   The headwords of marine engineering transparent compounds

backflow	eyebolts	outlet	shellfish
bedplate	feedback	overboard	shipbuilding
breakdown	flywheel	overflow	shutdown
bypass	freshwater	overhaul	sootblower
camshaft	gastroenteritis	overheating	standby
changeover	gastrointestinal	overlay	standstill
clockwise	handwashing	overload	thermometer
contouch	handwheel	override	throughput
copyright	inlet	overspeed	touchscreen
countdown	intake	overview	troubleshooting
crankcase	interlock	password	turbocharger
crankpin	intermediate	pipelines	upstream
crankshaft	keyboard	potentiometer	uptake
crankthrow	layout	screwdriver	upward
crosshead	login	seawater	warewashing
crosswise	millimeters	selfjector	whirlpool
datalogger	nameplate	servomotor	
dipstick	onboard	setpoint	
downstream	outbreak	setup	

For a Better Dictionary: Revisiting Ecolexicography as a New Paradigm

Xiqin Liu, *School of Foreign Languages / Research Center for Indian Ocean Island Countries, South China University of Technology, Guangzhou, China (flxqliu@scut.edu.cn)*

Jing Lyu, *School of Foreign Languages, South China Agricultural University, Guangzhou, China (Corresponding Author, jinglyu@qq.com)*
and
Dongping Zheng, *Department of Second Language Studies, University of Hawai'i at Mānoa, Honolulu, USA (zhengd@hawaii.edu)*

Abstract: Driven by practical conundrums that users often face in maximizing (e-)dictionaries as a companion resource, this article revisits and redefines ecolexicography as a new paradigm that situates compilers and users in a relational dynamic. Drawing insights from ecolinguistics and cognitive studies, it appeals for rethinking the compiler–user relationship and placing dictionaries in a distributed cognitive system. A multidimensional framework of ecolexicography is proposed, consisting of a micro-level and a macro-level. To the micro-level, both symbolic and cognitive dimensions are added: (1) the dictionary can be symbolically viewed as a semantic and semiotic ecology; (2) dialogicality should be highlighted as an essential aspect of e-dictionary compilation/ design, and distributed cognition can be emancipatory for rethinking dictionary use. The macro-level concerns the obligations of lexicographers as committed to three interrelated ecologies or ecosystems: language, socio-culture and nature. Transdisciplinary in nature, ecolexicography involves a holistic, systematic and integrative methodology to nourish lexicographical practice and research. Corpus-based Frame Analysis is introduced to identify ecologically destructive frames and ideologies so that the dictionary discourse could be reframed. The study upgrades our understanding of the ontological, epistemological and methodological aspects related to ecolexicography, serving as a call for philosophical reflections on metalexicography. It is also expected to create an opportunity for lexicographers to examine problems with (e-)dictionaries in a new light and dialogue about how to find solutions.

Keywords: E-DICTIONARY, LEARNER'S DICTIONARY, SEMANTIC ECOLOGY, SEMIOTIC ECOLOGY, ECOLINGUISTICS, ECOLEXICOGRAPHY, DIALOGICALITY, DISTRIBUTED COGNITION, SOCIO-CULTURE, CORPUS-BASED FRAME ANALYSIS, METALEXICOGRAPHY

Opsomming: Vir 'n verbeterde woordeboek: 'n Herbesoek aan die ekoleksikografie as nuwe paradigma. Voortgedryf deur praktiese probleme wat gebruikers dikwels ervaar in die maksimalisering van (e-)woordeboeke as 'n handboekhulpbron, word 'n herbesoek

aan die ekoleksikografie gebring en word dit geherdefinieer as nuwe paradigma wat samestellers en gebruikers in 'n relasionele dinamika posisioneer. Uit insigte wat verkry is uit die ekolinguistiek en kognitiewe studies word daar gevra om 'n herbesinning van die samesteller–gebruikers-verhouding en om woordeboeke in 'n verspreide kognitiewe stelsel te beskou. 'n Multidimensionele raamwerk van die ekoleksikografie, wat bestaan uit 'n mikro- en makrovlak, word voorgestel. Tot die mikrovlak word beide simboliese en kognitiewe dimensies gevoeg: (1) die woordeboek kan simbolies beskou word as semantiese en semiotiese ekologie; (2) diskoers moet beklemtoon word as 'n essensiële aspek van die samestelling/ontwerp van die e-woordeboek, en verspreide kognisie kan bevrydend wees vir die herbeskouing van woordeboekgebruik. Die makrovlak is gemoeid met die verpligting van leksikograwe wat verbind is tot drie ekologieë of ekostelsels wat onderling aan mekaar verbonde is: die taal, sosiokultuur en natuur. Die ekoleksikografie, transdissiplinêr van aard, behels 'n holistiese, sistematiese en integrerende metodologie om die leksikografiese praktyk en navorsing te voed. Korpusgebaseerde Raamanalise word gebruik om ekologies destruktiewe raamwerke en ideologieë te identifiseer sodat woordeboekdiskoers geherdefinieer kan word. Hierdie studie verbeter ons begrip van die ontologiese, epistemologiese en metodologiese aspekte wat verband hou met die ekoleksikografie, en ontlok filosofiese denke rakende die metaleksikografie. Daar word ook verwag dat dit 'n geleentheid vir leksikograwe sal bied om probleme rakende (e-)woordeboeke in 'n nuwe lig te ondersoek en vir gesprekvoering oor hoe om oplossings vir hierdie probleme te vind.

Sleutelwoorde: E-WOORDEBOEK, AANLEERDERSWOORDEBOEK, SEMANTIESE EKOLOGIE, SEMIOTIESE EKOLOGIE, EKOLINGUISTIEK, EKOLEKSIKOGRAFIE, DISKOERS, VERSPREIDE KOGNISIE, SOSIOKULTUUR, KORPUSGEBASEERDE RAAMANALISE, METALEKSIKOGRAFIE

1.   Introduction

Ecology refers to (the scientific study of) the relation of plants and living creatures to each other and to their surroundings. How organisms interact with one another and with their environment has become "a central question governing the survival and sustainability of human societies, cultures and languages" (Cronin 2017). Ecolinguistics (or ecological linguistics) investigates language in an ecological context. It explores the role of language in the human society and the ecosystem, and shows how linguistics can be used to address key ecological issues. This new branch of linguistics represents a turning point in language studies. Revolutionary in nature, it catalyzes the growth of many interdisciplinary fields of research. It distinguishes two positions for the ecological study of languages: one concerned with the relations between languages, and languages with the environment; the other investigating the interrelationships existing in a language (Albuquerque 2018). This distinction was first elaborated by Makkai (1993), who put forward the term "exoecological linguistics" for the former, and "endoecological linguistics" for the latter. They could be understood as the macrolevel and the micro-level in the framework of ecolinguistics.

Originating from lexicography and ecolinguistics, ecolexicography was first proposed by Sarmento (2000) as a part of applied linguistics, with a focus on addressing the effects and results that each lexeme brings to dictionary users. Sarmento (2005) argues that the main issue of ecolexicography is what the role of words is in our world and how a word can create, maintain or destroy a world. Many scholars (e.g. Hoey 2001; Tsunoda 2005) resonate with this viewpoint, stressing the importance of dictionaries as a tool of promoting linguistic diversity, socio-cultural harmony, and environmental sustainability. However, Sarmento (2000, 2002, 2005) holds that ecolexicography does not deal with the elaboration of ecology dictionaries or ecological terms. This perspective may be too limited as ecolexicography unavoidably faces the treatment of ecological vocabulary.

Albuquerque (2018) describes ecolexicography as a new discipline in lexicography and explores what it could contribute to pedagogical lexicography, especially in the analysis of dictionaries and the microstructure, and in producing teachers with a different worldview and in environmental education for students. He argues that eco-lexicography as a science should assist lexicographers to: develop a new way of looking at the world (the ecological vision of the world) and the words; realize the power of the words of a language for its speakers and for the world; offer ways to identify the ecological factors in language; and propose a new structure of article and definition (ibid.). He also points out that research on ecolexicography regarding these aspects is only at an embryonic stage, and it is necessary to lay a foundation for the ecolexicography approach that needs more researchers, research and projects. There is actually significant potential for (re)discovering important inroads or beneficial outcomes.

To breathe new life into this field, we have to re-examine the lexicographical products seriously, and rethink the cognitive and socio-cultural processes of dictionary compilation and use from a novel perspective. This article is expected to create an opportunity for lexicographers to dialogue about the problems they encounter with (e-)dictionaries and communicate how our ecolexicography proposal can shed light on the solutions it can provide.

2.   Rationale for revisiting ecolexicography

2.1   Practical problems: the necessity

Abundant literature (e.g. Hoey 2001; Tsunoda 2005) reveals that there are at least two kinds of problems with current dictionaries: anti-ecological language and destructive ideologies, and problematic (e-)dictionary design and use.

2.1.1   Anti-ecological language and destructive ideologies

Many dictionaries, including pedagogical dictionaries, are not ecologically ori-

ented and do not pay enough attention to users' awareness of the importance of environmental protection and sustainable development of human society or cultures (Wang 2003).

Tian et al. (2016) find that some examples in *The New Age English–Chinese Dictionary* (*NAECD*) fail to adopt a positive attitude toward ecology. Four tendencies of lexicographers dealing with biological and ecological lexemes were identified by Trampe (2001): (1) reification, i.e. treatment of certain living beings as things (goods of production or consumption), e.g. "cow" is a commodity; (2) use of euphemism (and other language mechanisms) to hide certain facts that may be regarded as violent for the consumer or general public, e.g. "pesticide" is replaced by "plant protection tool"; (3) defamation of traditional/ subsistence agriculture, which are generally labeled as being "unproductive", "expensive", etc.; (4) use of slogans and phraseological elements to convince the population that the destruction of the ecosystem is something natural/inevitable or even to disguise such destruction, affirming it as something good, e.g. "to create more wealth for all". These four tendencies alert lexicographers to the anti-ecological language of the world economic vision that is fragmented, increasingly alienating the human being from other species and nature (Albuquerque 2018).

Furthermore, anti-sociocultural ideologies are found in dictionaries. Tenorio (2000) claims that some definitions in *The Collins COBUILD English Language Dictionary* (*CCELD*) are inaccurate and biased in gender representation, and ignore changes in society. Hu et al. (2019) assert that *The Contemporary Chinese Dictionary* (*CCD*) portrays men as valuable social members while overlooking the value of women.

We found similar results (see Appendices I and II) after examining three of the "Big Five" dictionaries[1]: *Oxford Advanced Learner's Dictionary* (*OALD9*), *Longman Dictionary of Contemporary English* (*LDOCE5*) and *Collins COBUILD Advanced Learner's Dictionary* (*COBUILD8*). To achieve representativeness and generalizability of our data and the outcomes, we retrieved 30 random pages from each dictionary and all the linguistic data in those pages were collected to create a corpus. Each text was annotated and analyzed to disclose the ecologically (non)destructive frames in dictionaries. Frames (also called schemas) are schematizations of our experience and knowledge of the world (Fillmore 1985), and description of word meanings must be associated with cognitive frames in the reader's mind[2]. In our survey, we adapted and integrated corpus-based discourse analysis (Baker 2006) into frame analysis (Fillmore and Baker 2009; Lakoff 2014). The procedure of frame analysis (Blackmore and Holmes 2013) is to ask the following questions for a particular frame: What values does the frame embody? Is a response necessary? Can the frame be challenged? If so, how? Can (and should) a new frame be created?

We compared the frames represented by the headwords in the dictionaries and those represented by the same words in Corpus of Contemporary American English (COCA). In the end, we identified over 30 potentially destructive

instances (definitions and examples) from more than 30 entries of each of the three dictionaries. For instance, "He had abused his own daughter" and "The boy had been sexually abused" are used as illustrative examples in the entry of "abuse" in *OALD9* (see Part A in Appendix I). In total, we found that 23 themes were disharmoniously framed, and many of them were beyond the traditional lexicographic attention because it seems that the top seven themes we have identified (violence, animal, drug, possession, pollution, sea and alcohol) have not been fully discussed in lexicography (Lyu and Liu, in preparation). Destructive frames and ideologies (e.g. "Children are the target of sexual harassment", see Part A in Appendix II) seem to be prevalent, largely due to lexicographers' choice in this challenging age of the Anthropocene (ibid.).

2.1.2   Problematic (e-)dictionary design and use

Researchers find that many lexicographic e-products were developed with little influence from innovative theoretical suggestions and, as a result, current e-dictionaries often do not live up to the expectations of users and are misused by their users (cf. Gouws 2014). Many of them have problems including definition insufficiency and inaccuracy (Zhang 2015: 79-82), lack of customization (Liu, Zheng and Chen 2019), information overload (Gouws and Tarp 2017; Huang and Tarp 2021) and lack of education in dictionary use (e.g. Winestock and Jeong 2014). For instance, dictionaries integrated into English learning applications produced in China were found to suffer from deficiencies such as "inconsistent treatment of words and senses, data overload, difficult access, and inconvenient location of the pop-up window that displays the lexicographical items", which may "impact negatively on the learners' motivation and the learning process in general" (Huang and Tarp 2021). In the digital revolution, the way of displaying data in e-dictionaries must be redefined (Gouws 2014), and semiotic resources (e.g. color, typography, and navigation devices) should be properly employed according to the context (Liu 2015, 2017; Farina et al. 2019).

Underlying reasons for the above problems are complex, and some may be ontological and epistemological. At the fundamental level, many lexicographers, perhaps indulged in Western analytical thinking, still hold a fragmented view, rather than a systematic view of the components in a dictionary and its microstructure. There lacks an awareness that a dictionary, comparable to ecology, is characterized by complexity, holism, diversity and dynamicity. For example, the lack of e-dictionary customization and individualization is against the principle of ecological diversity and dynamicity. The technical transition from paper-based to electronic layout demands different cognitive attention and visual engagement. Users' individual and collective needs should be considered by designers. From an ecological perspective of language learning, even if a universal dictionary could be made, the users would tailor its use (especially those with a high degree of literacy and computer skills). So, dictionary design should try to allow users to adapt the product to their needs, goals and values,

to some extent (see Liu, Zheng and Chen 2019 for an example of varying types of motivation for smartphone dictionary use in China).

To make things even worse, the practice that one definition/example fits all, or lack of adaptability, may aggravate the problem of data overload. For instance, the word "pig" is defined as "An omnivorous domesticated hoofed mammal with sparse bristly hair and a flat snout for rooting in the soil, kept for its meat" in the *Lexico.com* (called *Oxford Dictionaries English* before 2019, https://www.lexico.com/definition/pig). This is a general dictionary (rather than a specialized one) and the definition is offered to users in general, but this is a very difficult technical definition. It is very likely that many users do not understand the difficult terminologies in the complicated explanation. Perhaps such information/data overload (Gouws and Tarp 2017), traceable to inconsideration of dictionary types and users, is against the principle of "ecological harmony" (cf. Zhou 2017). The idea that online dictionaries have unlimited space has furthered the often uncritical inclusion of too much data (Gouws and Tarp 2017).

In brief, the status quo highlights the importance of proper ontological and epistemological orientations for lexicography. With an ecological view, ecolexicography has the potential to offer a fresh set of theoretical-methodological contributions in dictionary research and compilation, especially in the proposal of a differentiated microstructure (Albuquerque 2018). Nevertheless, for systematic strategies to remedy the above problems, ecolexicography needs to be redefined as a new paradigm by drawing theoretical and methodological insights from related fields.

2.2   Theoretical underpinnings: the feasibility

2.2.1   Lexicographical theories

Three theories may shed new light on ecolexicography, the Communicative Theory of Lexicography (Yong and Peng 2007), the Function Theory of Lexicography (Bergenholtz and Nielsen 2006; Tarp 2007), and the Discourse Approach to Critical Lexicography (Chen 2019).

The first two theories are user-oriented and focus on the interactivity feature of dictionary compilation and use. The Communicative Theory of Lexicography views the dictionary as communication (instead of reference and text). Drawing insights from Systemic-Functional Linguistics (Halliday 1985), Yong and Peng (2007) assert that dictionary context encompasses three subcategories: field, mode and tenor. This communicative perspective inspires reconsideration of the interaction between dictionary compilers and users. According to the Function Theory of Lexicography (Bergenholtz and Nielsen 2006; Tarp 2007), dictionary functions are communication-orientated or cognition-orientated, and lexicographers must identify the relevant functions and select and present appropriate data so that the dictionary satisfies the needs of users in different situations.

Chen's (2019) Discourse Approach to Critical Lexicography, or Critical Lexicographical Discourse Studies (CLDS), offers both theoretical and methodological inspirations for ecolexicography. Responding to the call for lexicographers' social accountability, CLDS views the dictionary as discourse, and discourse is a three-tiered concept consisting of "a piece of text, an instance of discursive practice and an instance of social practice" (Fairclough 1992). To uncover the ideologies and power relations in dictionaries, CLDS analysts will first conduct an analysis of the dictionary as text, investigating, for example, the choice of vocabulary in explaining the meaning of a word, the choice of illustrative examples, and the order of senses (ibid.). Thereafter how the dictionary is produced, distributed and consumed will be examined, followed by a discussion of the social context in which the dictionary is produced and consumed (ibid.).

2.2.2   Ecolinguistic and cognitive theories

Two interrelated theoretical achievements in ecolinguistic and cognitive studies may offer nourishments for ecolexicography and help transform the discipline. The first is the "distributed language" and EDD (ecological, dialogical and distributed) theory (Van Lier 2002; Cowley 2011; Linell 2009, 2013; Zheng 2012; Steffensen 2015), and the second is Steffensen and Fill's (2014) redefinition of ecolinguistics by identifying the four ways in which the ecology of language is conceptualized.

Distributed language theory means that language is not an independent symbolic system used by individuals for communication but rather an array of behaviors that constitute human interaction. Language perception occurs in a context of activity and interactivity (Van Lier 2002). Permeating the collective, individual and affective life of living beings, language is a profoundly distributed, multi-centric activity as a part of our ecology, and it gives us an extended ecology in which our co-ordination is saturated by values and norms that are derived from our sociocultural environment (Cowley 2011). In brief, language (or language use) is ecological, dialogical (linked to others) and distributed (rather than located to any single place, such as the speaker's brain) (Zheng 2012).

In applied linguistics, Van Lier (2002) might be the first to have introduced an ecological perspective to language education. The ecological view has inspired a rethink of language and language acquisition/cognition from a sociocultural perspective and boosted the development of such emerging theories as "the Complexity Theory" (see Larsen-Freeman 2011). Ecolinguists redefine language by dividing it into two different consensual domains: (1) first-order languaging (linguistic actions and activities in the communication); (2) second-order sociocultural inscriptions and norms (Kravchenko 2009). Following this theoretical vein, Zheng (2012) proposed her ecological view of language learning and use which highlights the dialogicality and distributed cognition of participants in communication. Distributed cognition means that cognition is spread in and reliant on different contexts. Traditional cognition is redefined as

an activity "distributed" in the physical and socio-cultural environment. In cultural ecologies, resources like a dictionary can link people in practices that enable the accomplishment of tasks.

In ecological terms, agents' languaging behaviors are caused not by stimuli but the affordances, opportunities for action and coaction motivated by the ecosocial environments (Zheng et al. 2012). Language is embodied (not merely abstractly procedural), embedded (shaping and shaped by social systems in a cultural world), enacted (living in or realized in and through action), extended, situated, and multi-scalar (existing on different time-scales) (Cowley 2011; Linell 2013).

Based on the communication models of semiotic activity by Zheng (2012) and Linell (2009), we build an ecological model of lexicographical interaction (see Figure 1).
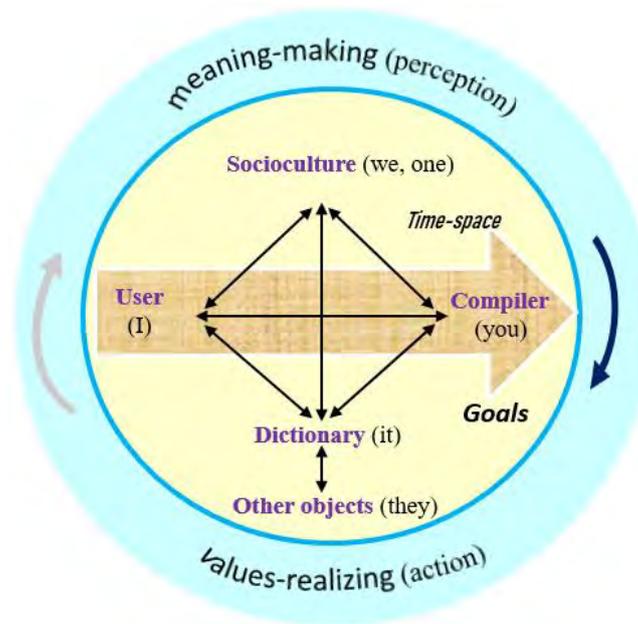
Figure 1:   An ecological model of lexicographical interaction: dialogicality and distributed cognition

In the outer layer of the model, there are two concepts from ecological psychology, meaning-making (perception system) and values-realizing (action system). Values-realizing means that an individual agent makes "a conscious choice among multiple values at play in any given moment of action and interaction" (Zheng et al. 2012). It is values that "guide the selection and revision of goals across

diverse time-space scales, under which the sociocultural norm 'we' (laws or rules of phonology, syntax, or semantics) are nested" (Zheng 2012). There are interactions among the dictionary user (I), compiler (you), sociocultural norm (we), dictionary (it) and other objects (they) in the real world or virtual space.

Based on the model (Figure 1), the relationship between the dictionary (it) and its user (I) should be rethought. The dictionary should be a friend that is always there, so faithful, helpful and thoughtful. This means that it should have such qualities as accuracy, functionality and adaptability. In addition, the interaction between the dictionary (it) and the other objects (they) in the physical environment is also meaningful. To improve its adaptability and customization, an e-dictionary is often embedded in or fused with the interfaces of learning activities like those of reading or writing software. Meaning-making and values-realizing are in the cycle of perception and action involving dictionary compilation and use.

Another illuminating insight that ecolexicography can gain from ecolinguistics contributes to an upgraded understanding of its overall framework. Steffensen and Fill (2014) point out four ways the language ecology has been conceptualized as a symbolic ecology, a cognitive ecology, a natural ecology, and a sociocultural ecology. Similarly, in terms of ecolexicography, a symbolic ecology can be understood as the semantic and semiotic ecology in a dictionary. A cognitive ecology of lexicography involves dictionary compilation/design as dialogism and dictionary use as distributed cognition. The two constitute the microlevel of ecolexicography. At the macrolevel, ecolexicography should be committed to serving the linguistic, natural and sociocultural ecologies. The differentiation (and complementarity) between the microlevel and the macrolevel of ecolexicography mirrors the exoecological vs. endoecological division in ecolinguistics.

The endoecological position or the microlevel of ecolexicography, an obvious lacuna in literature, needs to be delineated to form a complete framework. This article aims to take a small step toward addressing the gap by revisiting ecolexicography as a new paradigm.

3.   Ecolexicography at the micro-level

3.1   The semantic and semiotic ecology in a dictionary

Some scholars (e.g. Liu 2015) hold that the dictionary as a complex system can be symbolically compared to an ecology in two senses, semantic and semiotic.

First and foremost, the complicated semantic system of a dictionary is comparable to an ecology where meaning is like energy. Meaning flows and expands (Liu 2017), just as energy flows and circulates. As an ecological system, an e-dictionary is even more open and dynamic than a paper dictionary. There are interactions among the diverse members in this ecology, including

cooperation and competition. Its components or communities are conceptually linked together as an integrated whole in a hierarchy. This organic whole is served by the dictionary macrostructure, mediostructure and microstructure as well as other information components that are themselves reciprocally conditioned. The macrostructure is the form and size of a dictionary, the mediostructure refers to its system of cross-referencing which can create textual cohesion, and the microstructure means its lexical entries or articles. Figure 2 roughly illustrates a pyramid of the dictionary semantic(-functional) ecology.
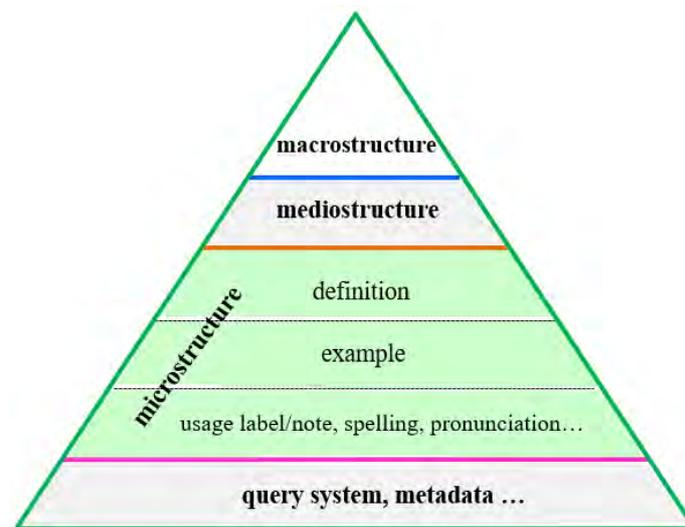
Figure 2:   Pyramid of the semantic(-functional) ecology of a dictionary

The macrostructure of a dictionary is like an overall guide or a head that leads the whole semantic ecology on the top. The mediostructure is a network under the macrostructure. In the digital era, cross-references between words by hyperlinks easily connect entries and reinforce the mediostructure. The microstructure is the main body of a dictionary where definition acts as the core of meaning representation, playing a key role in stating or explaining the meaning of a word or phrase. A definition is often complemented by the illustrative examples ("examples" hereafter) under the same sense. Examples can reinforce meaning explanations, illustrate collocations and colligations, and contextualize for cultural, stylistic and pragmatic implications (Xu 2009: 12, 26-29). Many examples are transformed (i.e., simplified for children) from authentic sentences to meet the particular purposes of a dictionary, a lexicographical process like crop improvement in the biological ecology. Other microstructure components, such as spelling, pronunciation, usage notes and labels, also participate in the co-

construction of meaning, serving behind as a guide to the microstructure. Other information components (query system, metadata etc.) are backgrounded on the bottom of the ecology. They act as the supporting system.

Furthermore, the semiotic system in a dictionary can be regarded as an ecology that is increasingly diversified in the digital era. In an e-dictionary, there are three major categories of multimodal meaning-making devices: written language, audio presentation of the verbal elements, including human voice recordings and synthesized speech, and other devices (Lew 2010), like pictures, silent animations, video clips, hyperlinks, floating tips and typography. Based on Multimodal Discourse Analysis (Kress and Van Leeuwen 2006) the choice of semiotic modes, and the cooperation and interaction among different modes are important for the dictionary ecology (Liu 2015).

If the semiotic ecology is examined hierarchically from a Systemic-Functional perspective, it can be stratified into three levels by following Rossi and Sindoni (2017): (1) semiotic systems (i.e. ideational, interpersonal and textual metafunctions or meaning potentials); (2) semiotic resources (i.e. instance or realization); (3) semiotic components (i.e. elements that can be unpacked from a resource and that concur to the instantiation of texts). They represent different levels of abstractions. Figure 3 illustrates the different strata proposed in our analysis.
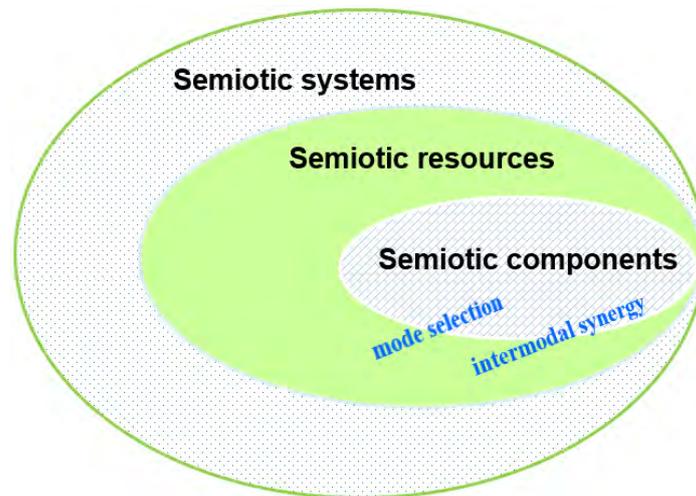
Figure 3:   Stratification of the semiotic ecology of a dictionary

In brief, the semantic ecology is organized holistically by the synergy of multimodal devices, so that the whole is more than the sum of its parts.

3.2   The dialogicality in (e-)dictionary compilation/design

Dialogicality means the dynamic abilities of human beings to take part in inter-actions with others and with sociocultural contexts as well as physical envi-ronments (Linell 2009: 368). Meaning or sense is co-constructed by dictionary compilers/designers and users, and it is not local. In the era of Web 2.0 and Web 3.0, there are more chances for them to have dialogues to make meaning/sense. Problems such as lack of customization and information overload in dic-tionaries (Gouws and Tarp 2017) can be alleviated with a dialogical perspective.

Compared with paper dictionaries, e-dictionaries provide users with more chances of participation and interaction, facilitating compiler–user or even user–user dialogues and greater flexibility in use (cf. Liu 2017). Many e-diction-aries invite users to contribute entries or make comments on them (Granger 2012). They afford user customization. To take a previous version of *Jinshan Ciba Eng-lish Dictionaries* (*iCIBA*) as an example, classified information was provided and its users could choose the type of examples they wanted. There were also but-tons users could click to report a wrong example to the designer, praise a good one "in public" and save a useful one for his or her own use. Figure 4 is a screenshot of its entry of "ecological" (captured on Jan. 20, 2018)[3]. We have added English translations for its customization and interaction buttons.

Figure 4:   Customization and interaction buttons on the interface of *iCIBA*

It is advisable for e-dictionary designers to consider the user's ecological niche and allow co-selectivity and co-creation of meaning and value[4]. Learners' goals and needs could be scaffolded and transformed by design. This suggestion is broadly informed by distributed language, in that first-order dynamic action should be at the fore of second-order static prescription (Cowley 2011). The

design based on the traditional concept that, on one hand, there is an objective, absolute authority of dictionary meaning, and on the other, there are users who use this absolute value-free tool, should be rethought (Liu, Zheng and Chen 2019). From the ecological psychological perspective (Gibson 1979), dictionaries can be considered having affordances for certain actions, such as for supporting enjoyment of reading, and for clarifying a statement. The ways in which actions connect dictionaries and users should result in changes in both artifacts and the agent (cf. Zheng 2012). Therefore, how dictionaries are designed can have a direct influence on learner behaviors.

Besides having more dialogues between dictionary compilers/designers and users, the ecology of an e-dictionary is filled with more different voices than that of a paper dictionary. Pop-up windows, for example, are used for projecting the voices of advertisers. With the social force of marketization, dictionary companies have to attract funding from advertisements to maintain the dictionary. Creating a more heteroglossic and noisy atmosphere, advertisements add to the complexity of the semantic ecology and may often distract users' attention in their cognitive processes. By heteroglossia, we mean a diversity or hybridity of voices and styles of discourse in the dictionary ecology as an extension from lexicographical dialogism[5].

3.3   The distributed cognition in (e-)dictionary use

In ecolexicography, "distributed cognition" can be understood in both narrow and broad senses. In a narrow sense, the page layout of e-dictionaries usually looks less cluttered and the user's cognition involved in consultations is not restricted in a fixed manner. In a broad sense, cognition is distributed over different systems, such as brain, body, computers, instruments, aspects of the environment at large (Steffensen 2015).

The digital revolution of dictionaries "has removed constraints on size and format, paving the way for multi-faceted, flexible and rich representations of word meaning and use" (Fellbaum 2014). In densely printed pages of text, reading is often linear and strictly coded (Van Leeuwen 2005: 204). The one-time display in a fixed order might leave the users in a passive state of reception (Liu 2017). Large bands of space can be found in contemporary designs, suggesting the lightness of the reading experience. Spatial resource competition is less fierce.

In an e-dictionary, individual examples are often placed in separate paragraphs, and this makes them more readable than those densely printed in paper dictionaries. A distinctive type of vertical composition for examples in smartphone or tablet dictionary applications (apps) has been identified from *Fayu Zhushou French Dictionaries* (a most popular French dictionary app in China), where elements are mostly placed into equally sized tiles which could be swiped across to see more (Liu 2017). This way of organizing information allows contents to be textually linked as choices of the same order since tiles of the

same size also achieve textual linking or rhyming, alongside that accomplished by color and fonts (Zhang et al. 2015). This creates visual harmony in the dictionary ecology, decreasing the difficulty of reading on a small screen. Also, there is added convenience of the ability to zoom in and out, not to mention the possibility to blend with other assistive technology to aid the visually impaired and those with poor eyesight. Users can make use of the convenient features of copying entries and exporting to other applications, or sending selected text through messenger applications.

E-dictionary information can be presented in an array of interlinked web pages and media networks, enabling e-dictionary users to navigate and choose their own pathways through this semantic ecology. Users can change dictionary settings (like interface style[6]), and make bookmarks, tailoring the use according to their own needs (ibid.). For instance, as shown by Figure 4, the "Learn" button at the end of each example can be clicked to start a timed activity of memorizing an example, inviting the user to put the disordered words of an example sentence in good order. In the online *Longman Dictionary of Contemporary English* (*LDCE*), a collocation in its examples is highlighted with an underline and boldface font when the user's mouse cursor hovers over it (Liu 2017). During such human-computer interactions, a solid line emerges, giving readers a sense of the formation of the bundle/collocation (ibid.). This certainly involves distributed cognition.

In addition, because cognition is distributed across different places and contexts, effective lexicographic solutions should be suited to the needs of a particular user in a particular situation (Lew 2012), especially outside the classroom. For example, smartphone dictionary apps, due to the nature of portability, can work as flashcards for learners to carry with them. The smartphone can be used to scan an unfamiliar word for its meaning or translation and the user can add it to a wordlist for learning. Learning becomes more contextualized and meaningful when tied to learners' lives outside academia, and mobile devices help achieve that goal (Godwin-Jones 2011). Dictionary use, in turn, can transform language learning behavior as distributed cognition.

Furthermore, education in e-dictionary use can be provided from an ecological-dialogical perspective. This means that learning happens in an ecology with interactions between dictionary users and designers, between learners and teachers, and among learners (Liu, Zheng and Chen 2019). Dictionary use can be a result of a meaningful situated activity in which users need to consult a dictionary to understand meaning (Zheng 2012). Action-based activities can be better realized if instant support can be provided in a specific situation (Zheng et al. 2015) with a smartphone dictionary. Also, situated and action-based activities can integrate with dictionary user training in a natural way.

In brief, with the proper use of semiotic resources, e-dictionaries can facilitate distributed cognition effectively. As a result, the role of the dictionary user changes from a passive receiver of meanings to an active explorer of senses. Users are unable to maximize a dictionary as a companion resource because

dictionaries are conceptualized as an object that supplies predefined meanings. This article explores rethinking that a dictionary is a relational component of a distributed cognitive system along with users and compliers. Thus, the use of a dictionary in this new thinking helps make meaning with the distributed system.

4.   Ecolexicography at the macro-level

4.1   Lexicography and the ecology of language

A dictionary is a good tool for outsiders (non-native speakers) to learn the language, and for insiders to document their language (Lee 2017: 5). The dictionary community is situated in its language habitat, and provides a prerequisite for the recording and development of natural language life. Etymology dictionaries play a most important role in documenting the minority languages which could be endangered because these languages contain and offer unique experiences of nature and knowledge, which have to be saved for future generations — especially in the sense of sustainability (cf. Bang and Trampe 2014). An average of 6 languages are disappearing from this world every year, and 1,531 languages among 7,102 are classified as threatened or shifting (Lee 2017: iii), so there is much work to do to revitalize endangered languages. If possible, a holistic approach (Tsunoda 2005: 231-233) may be the best way to document one language, covering various aspects of a language, including phonology, morphology, syntax, discourse, semantics, and vocabulary.

Methodologically, compiling dictionaries for endangered languages is different from that for languages that are not under the threat of extinction (ibid.). For the former, lexicographers should highlight the changes in the manner in which it is used, the reduction of the number of its different registers, as well as changes and simplifications in its structure, and lexical composition, and semantic changes in its lexicon, with all this resulting from a linguistically-oriented endangerment of its traditional form (Wurm 2007). At the same time, they should be aware of the sociolinguistic aspects, like the declining use of the language by shrinking numbers of its speakers, and the reasons, and circumstances of such events (ibid.).

Although some dictionaries documenting endangered languages are products of individual or community efforts, like *Buk Bang Sinda* (*Bidayuh–Malay–English Dictionary*), most of them result from a "top-down" process. Lexicography can be regarded as a part of the language planning of state agencies. Language planning was first introduced in 1959 by the ecolinguistic forerunner, Einar Haugen, and the subject has become increasingly important as awareness of the socio-political nature of language choices in multilingual/ multi-dialectal communities has grown (Jones 2015: xiii). The ideologies underlying language planning strategies are often, at least partly, attributable to what has been described as language policy (ibid.). The first step in saving dying

languages is to persuade the world's majorities to provide opportunities for the minorities among them to speak with their own voices. Compiling dictionaries of minority languages may need teamwork among lexicographers, sociolinguists, ethnographers and anthropologists. An ecological perspective would be preserving not only languages but also the social group. Without people and community, what would language be for?

Bosch and Griesel (2020) proposed an innovative way of documenting and preserving nine African languages in a digital lexical database, the African Wordnet. They claim that such a database becomes a useful resource for natural language processing, consolidating dispersed indigenous knowledge collected from a variety of sources in a digitized hierarchical wordnet structure.

4.2   Lexicography and the ecology of socio-culture

It is generally acknowledged that dictionaries are not value-free representations of languages and the world. Illustrative examples, for instance, are imbued with lexicographic intentions and "constitute a repository of the common values and interests of the society whose language is described" (Béjoint 2010: 202). Dictionaries should convey ideologies in such a way as to promote the positive development of human society, including peace, justice, equality and sustainability. Unavoidably, social learning must be moral learning (Hodges and Baron 2007), and values are not properties of persons or objects, but relationships and the demands that the ecosystem places on those relationships (Zheng 2012). As frames are mental structures that shape the way we see the world and we know frames through language, morally based framing[7] is everybody's job (Lakoff 2014: 116), including lexicographers.

In the field of critical lexicography, scholars have examined such issues as gender (e.g. Hoey 2001; Moon 2014), racism and religion (e.g. Willinsky 1994; Ogilvie 2013), and politics and class (e.g. Ezquerra 1995). Previous studies (e.g. Benson 2002; Hornscheidt 2008) reveal how imperialism, racism and colonialism are naturalized in dominant monolingual dictionaries, such as *the Oxford English Dictionary*. Chen (2015, 2017) argues that bilingual lexicography is a complex site of ideological struggle and recontextualization of lexicographical discourse across cultures and contexts, resulting in the transformation and transfer of meaning. Recontextualizers of *A New English–Chinese Dictionary* (*NECD*), for instance, attempt to de-politicize the words and examples in the source dictionary by using such discursive strategies as deleting, replacing, and re-signifying (ibid.). When we successfully reframe dictionary discourse, we change the way the user sees the world. Because language activates frames, new language is required for new frames (Lakoff 2014: 15).

Many examples in Big Five dictionaries are found to embody ideas and values which are biased or politically wrong. For instance, *COBUILD8* presents "Possession of cannabis will no longer be an arrestable offence" for the entry of "arrestable", "I started smoking grass when I was about sixteen" for "grass" (= mari-

juana), and "Up to two thirds of 14 to 16 year olds admit to buying drink illegally ..." for "admit" (see Part A in Appendix II). These examples, scattered in different entries, could co-build a harmful frame of drugs in the user's mind: Using drugs is a good experience; People can legally be drug abusers. Furthermore, a 2020 slang phrase, "Funny mud pee" (Go fuck yourself), included in the crowdsourced *Urban Dictionary* (https://www.urbandictionary.com/define.php?term=Funny%20mud%20pee), is an example of racism recontextualized from the Chinese-speaking to English-speaking world. It is a Chinglish curse created by Chinese social media users posting on Twitter extensively amid the global outbreak of COVID-19 in response to those tweets labeling COVID-19 as "Chinese virus" or "Wuhan virus". The lexeme is used in the media as something bad and must be eliminated or annotated properly to avoid harm or destruction of the global community.

Information in dictionaries should be selected and presented in such a way as to respect various cultures. In *LDCE*, the word "pig" is defined as "a farm animal with short legs, a fat body, and a curved tail. Pigs are kept for their meat, which includes pork, bacon, and ham" (https://www.ldoceonline.com/dictionary/pig). From a socio-cultural perspective, this definition is incomplete since it is not true that pigs are kept for their meat by all communities. In some cultures, pigs are kept as pets. Muslims don't eat pork and they would feel uncomfortable when reading such a definition.

In sum, the relationship between ideology and dictionary compilation is not a new topic in lexicographical studies, but most previous research has explored the topic from the perspective of dictionary compilers or designers. Attempts should be made to study the relationship from the perspective of the response of users regarding monolingual and bilingual/multilingual dictionaries. Users have the power to accept or reject a dictionary or a definition or example in an entry that is ideologically similar to or different from their own thinking, and ideology-oriented dictionaries can only realize their purpose in the right place at the right time (cf. Li 2012). After all, dictionary use ― as a part of languaging ― can promote "individualized values-realization" (Zheng 2012). That is a new research orientation that macro-ecolexicography can explore.

4.3   Lexicography and the ecology of nature

Language can inspire us to destroy or protect the ecosystems that life depends on. For example, the language of advertising can encourage us to desire unnecessary and environmentally damaging products, while nature writing can inspire respect for the natural world (Stibbe 2015: 174). Dictionaries are committed to the task of ecosystem protection, and help address such overarching ecological challenges as biodiversity loss, food security, climate change, water depletion, energy security, and chemical contamination. At the very beginning of dictionary design, the headwords which are closely related to the ecosystem can be selected in a separate list and given special attention. *Wordsmyth Kids Dictionary*

(*WILD*, a popular children's dictionary) offered to guide children to explore words about the world, putting them in two modules: those about the city and those about nature. It seems that *WILD* embodies such a destructive frame: It is normal that urban residents stay away from nature since they don't belong to it. In other words, an implied contrast between city and nature may convey ideological ideas against human-nature oneness.

In writing definitions, lexicographers can implicitly or explicitly reinforce the users' awareness of environmental protection. This educational function of dictionaries can never be underscored enough. Take the word "ermine" as an example, in *Cambridge English Dictionary (CED)*, it is defined as "expensive white fur with black spots that is the winter fur of the stoat (= a small mammal) and is used to decorate formal clothes worn by kings, queens, judges, etc." (https://dictionary.cambridge.org/dictionary/english/ermine). This definition has connotations of the merciless killing of the animal, and it implies the glory of wearing ermine clothes with a focus on the economic benefits from ermine trade. There are similar definitions in *LDOCE5*, e.g. "the skin or fur of some types of seal, used for making leather or clothes" for the entry of "sealskin", and "strong soft leather made from the skin of a deer or goat" for "buckskin" (see Part B in Appendix II). This "reification" (Trampe 2001: 1) of animals may constitute biased representation. The embedding of humans in the larger systems that support life is forgotten or overlooked (Stibbe 2014: 585), making it possible to treat animals as commodities at the service of human needs (Fusari 2018). This makes the dictionary evade its function of environmental protection, similar to or perhaps even worse than "animal erasure" (Stibbe 2015: 155). Lexicographers should be aware that dichotomic representations opposing animals to humans are deeply rooted in language, and make use of lexical or grammatical devices to create public consensus in favor of effective conservation of biodiversity (cf. Fusari 2018).

While choosing examples for an entry from the corpus, lexicographers could zoom in on the texts and discourse of ecological importance, and then choose and adapt examples carefully. In *Merriam-Webster's Learner's Dictionary* (*MLD*), examples are creatively designed for the first sense of the headword "nature" (http://www.learnersdictionary.com/definition/nature). As Figure 5 shows, the six examples are coherently organized and the second ("She is a real nature lover.") aims to cultivate love of nature in particular. The way the examples are ordered is carefully chosen. As their linguistic difficulty increases, they loom progressively into an integrated discourse that communicates important educational messages: nature is beautiful, it deserves our love and study, we can explore it (including its color) by taking photos, and we should conserve nature. The examples construct a harmonious semantic ecology, embodying human-nature oneness with schematic experience and knowledge[8].

1 [noncount] *also* **Nature**
a : the physical world and everything in it (such as plants, animals, mountains, oceans, stars, etc.) that is not made by people

- the beauty of *nature*
- She is a real *nature* lover. = She really loves *nature*. [=she loves to spend time outdoors]
- He devoted himself to the study of *nature.*
- That is a color not found in *nature.*
- *nature* photography
- *nature* conservation

[-] hide examples

Figure 5:   Screenshot of an entry in *MLD*

Compilers of ecologically-minded dictionaries can construct subcorpora of texts about the ecosystem and environmental protection, like personal virtual corpora based on the *Wikipedia* Corpus on the platform of Brigham Young University (https://corpus.byu.edu/wiki/). Based on such ecologically-oriented corpora, lexical databases can be constructed and shared by world lexicographers. EcoLexicon is such a terminological knowledge base on the environment (http://ecolexicon.ugr.es) with terms in six languages: English, French, German, Modern Greek, Russian, and Spanish. It is the practical application of Frame-based Terminology to configure specialized domains on the basis of definitional templates and create situated representations for specialized knowledge concepts. We should develop an awareness of green lexicographical technology and improve ecological efficiency in the ecosystem of lexicography. Then the dictionary discourse can be reframed effectively, eliminating destructive frames systematically. For instance, the frame of "Animals are resources for human abuse" (see Part B in Appendix II) seems prevalent and deeply entrenched in "Big Five" dictionaries, and this systematic problem could be solved with the help of ecologically-oriented databases.

Definitions and examples should try to reflect the reality and dynamism of bio-ecology. Take the headword "romaine" as an example, it is defined as "a type of bitter-tasting lettuce with long leaves" in *LDCE* (https://www. ldoceonline.com/dictionary/romaine). But in actual fact, this vegetable sold in the supermarket now is not bitter at all, perhaps as a result of long-time crop improvement. So, this definition may either be against the reality of bio-ecology or fail to reflect its changes.

Proper notes or labels concerning ecosystem protection could be added to an entry of ecological importance. In general-purpose dictionaries, encyclopedic information could be added by referring to specialized dictionaries on

ecology, such as *Dictionary of Environment and Ecology*. This integration of linguistic and non-linguistic information corresponds with the holistic view of meaning and the functional theory of lexicography (Tarp 2007).

Serious ecological destruction has already occurred and more would be inevitable, so resilience to further environmental changes is necessary for finding new forms of society (Stibbe 2015: 15). This resilience can be properly embodied in lexicography, which connects the natural and social layers around dictionaries.

5.   Discussion

5.1   A unified framework of ecolexicography

Based on Steffensen and Fill's (2014) conceptualization of the language ecology as a symbolic, cognitive, sociocultural and natural ecology, we have identified five new dimensions that an ecological perspective can add to lexicography: symbolism, cognition, language, socio-culture and nature. These dimensions fall into two levels of ecolexicography: microlevel and macrolevel. The first two dimensions (symbolism and cognition) constitute the microlevel of ecolexicography. The other three dimensions (language, socio-culture and nature) form the macrolevel. The two levels can be regarded as two domains of the area, micro-ecolexicography and macro-ecolexicography. We tentatively propose a framework of ecolexicography unifying the two levels or domains (see Figure 6).
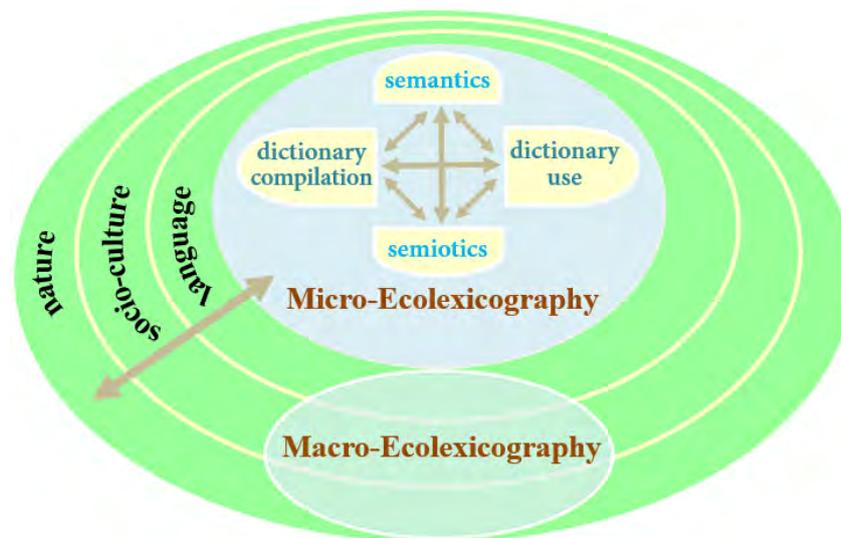
Figure 6:   A multidimensional framework of ecolexicography

We may draw insights from multimodal discourse analysis to interpret the layout of Figure 6. According to the information value principles for visual composition proposed by Kress and Van Leeuwen (2006: 197), it is fundamentally a structure of "Center and Margin". In this model, micro-ecolexicography is the Center and it forms the nucleus of the space. Three outer layers (the Margins), language, socio-culture and nature, "wrap" or contextualize the cognitive processes of dictionary compilation/design and use as well as the dictionary itself. These layers create a gradual and graded distinction between Center and Margin. In micro-ecolexicography, dictionary compilation is on the left denoting the Given (i.e. old information) while dictionary use is on the right denoting the New (i.e. new information). Semantics is on the top representing the Ideal while semiotics is on the bottom representing the Real. For something ideal means that it is presented as the idealized or generalized essence of the information while the Real presents more specific information (e.g. details), more down-to-earth or practical information (ibid.: 186-187).

The framework represents an interactive ecology, and it refreshes our understanding of the major tenets of ecolinguistics through a lexicographical lens. Micro-ecolexicography is the origin and prime mover of communication. The double-headed arrows in Figure 6 indicate dialogues and interactions among communicative participants or affordances as well as ecologies, symbolizing dynamism or circulation of the ecosystem. There are underlying linkages and interactions between every two of the four "layers" (dictionary, language, socio-culture and nature).

Micro-ecolexicography may focus on e-dictionary design and use in the digital era while macro-ecolexicography highlights dictionaries' educational function and lexicographers' commitment to three interrelated ecologies (nature, socio-cultural and language) and can be attentive to various types of dictionaries.

5.2   Rethinking the methodology of ecolexicography

The methodology of ecolexicography should be reconsidered as a new paradigm. Ecolexicographic thinking is concerned with complex systems and diverse situations. It is arguable to place dictionaries within a distributed cognitive system and view them as such a system. Holistic, systematic and integrative methodology is essential to ecolexicographical practice and research.

Micro-ecolexicography could benefit a lot from the following methodologies: (1) multimodal discourse analysis for building a semiotic and semantic ecology of the dictionary; (2) information technology (e.g. data mining for tracking user feedback) and transdisciplinary approaches to user research for enhancing dialogicality and distributed cognition.

User research should be a key theme of micro-ecolexicography. It can be done before a dictionary is compiled so that preventive measures could be taken by its designer. A statistical method widely used in psychometrics, latent

class modeling, should be a useful tool for investigating user intentions and attitudes. It identifies the underlying or invisible subgroups/categories (e.g. motivation) in the population, and can be introduced to large-scale surveys (see Liu, Zheng and Chen 2019 for an example). Surveys with latent class modeling could offer more important pointers for dictionary customization than traditional surveys that are generally based on visible subgroups (e.g. gender) of dictionary users.

However, when investigating socially sensitive attitudes like racial prejudice through a survey or interview, one should be aware that people are often motivated to self-report unprejudiced and egalitarian beliefs. To bypass social-desirability bias, an experimental paradigm of cognitive psychology, the Implicit Association Test (IAT), is recommended. As an influential measure of people's unconscious attitudes, IAT is less subject to deliberate control and potential distortion than interviews or focus groups. Such interdisciplinary methods have advantages over traditional approaches in exploring underlying values and frames in the minds of dictionary users and even dictionary designers.

Macro-ecolexicography could be nourished by: (1) Critical Lexicographical Discourse Studies for globally critiquing the dictionary discourse; (2) Corpus-based Frame Analysis (CFA) is recommended for revealing ecologically (non)destructive ideologies and frames in dictionaries specifically (cf. Lyu and Liu, in preparation); (3) ways of reframing the dictionary discourse.

Chen's (2019) Critical Lexicographical Discourse Studies (CLDS) not only offers a theoretical rationale for revisiting ecolexicography (see Section 2.2.1), but also specifies how to critique the dictionary discourse. With a focus on inter-discursivity, it compares the dictionary discourse with other types of texts. According to Chen (2019), analysis of the order of discourse is first done to disclose the latent social rules that govern the production of discourse. Then an interactional analysis is made, which consists of interdiscursive analysis, and linguistic/semiotic analysis (e.g. identifying how the lexicographic discourse is interdiscursively related to other discourses and genres, and analyzing lexis and clauses). Such an integrated analysis of textual and social structures helps uncover the internal relationship between them, with the situated contexts of dictionaries taken into consideration (ibid.).

If CLDS provides a macroscopic view for reframing the dictionary discourse, frame analysis, a common tool for critiquing or promoting ecological discourse, is perhaps more microscopically oriented. Corpus-based Frame Analysis could be more reliable than traditional frame analysis since it exploits objective evidence of corpus data for comparative analysis of ideologies and frames. Examining the semantic roles and their interrelations within a text or across texts via corpus analysis can reveal the ideologies and cognitive frames behind them. We also draw insights from previous scholars (Blackmore and Holmes 2013: 42; Stibbe 2015: 46-61) who have introduced the social-values-oriented frames to ecolinguistic research from the perspectives of cognitive semantics and discourse studies.

Corpus-based Frame Analysis consists of four main steps (see a more detailed illustration in Lyu and Liu, in preparation):

(1) randomly sampling definitions and/or examples from a dictionary and building them into a mini-corpus of dictionary discourse (D-corpus in short);

(2) identifying destructive discourse and analyzing frames behind the sample data (see Endnote 2 for the procedure of frame analysis), examining their relations and classifying them into a hierarchical network when necessary;

(3) using keywords to extract relevant discourse and analyze frames from an authoritative (and presumably balanced) corpus (namely B-corpus) or the corpus the dictionary claims to be based on (if available), and comparing the B-corpus frames with the D-corpus frames;

(4) based on the features and distribution of destructive frames, exploring the possible reasons for their existence with contextual factors considered, and rethinking its social accountabilities and possible solutions.

A keyword may have numerous frames in the B-corpus, and only those reflected by the most frequent collocations are considered. For example, for the patterns "A motivate B to do" and "A lead B to do" in the dictionary, the most frequent collocations in COCA turn out to be: factors motivate/lead somebody (sb) to do something (sth); sth motivate students to do sth; sth lead people to do sth; sth motivate sb to develop sth; sth lead sb to believe sth else. As Part A of Appendix II shows, *LDOCE5*, backdropped against COCA, creates a spouse killing frame through these examples: "We may never know what motivated him to kill his wife" and "What led him to kill his wife?". This is a biased representation of both real life and language use.

Destructive frames have their own linguistic and distributional features as they may be reflected by different (numbers of) definitions and examples in the dictionary discourse. The reasons for their existence can be traced to these features. Some frames are widely distributed and may form a complicated network. According to Appendix II, the frame of animal abuse seems far more complex (divisible into four subtypes in this case) and perhaps more severe than the frame of plant abuse, so the former may deserve more attention and a systematic solution is necessary.

The ultimate purpose of CFA is to reframe the dictionary discourse to avoid biased representations of reality. Dictionaries are supposed to capture the most typically shared values and ethics of a community to represent them in the definitions and illustrative examples (see Figure 5 for an example). At least, such problems as Trampe (2001) identified (e.g. reification, defamation, disguise) should be rectified. We'd like to recommend five ways of reframing the dictionary discourse: warning, commenting, refining, questioning and neutralizing.

The most straightforward way of reframing is to give a warning against

immoral and illegal practices. For instance, a definition of "crocodile" in *LDOCE5* ("the skin of this animal, used for making things such as shoes", see Part B in Appendix I) may be inadequate because it seems to ignore that crocodile is a species at risk of extinction. This definition could be refined to reveal the fact, or a note/warning could be added (e.g. "The crocodile is an endangered animal and should not be killed at will for profits). Another direct way is to enhance existing linguistic data by commenting on immoral values and improper behaviors. For instance, one could present "Experimentation with cannabis is illegal", rather than "experimentation with cannabis" as found in *LDOCE5* (see Part B of Appendix I).

The third way of reframing, refining, means changing the current definitions or examples moderately, often by adding modifiers or other descriptions of details. Take the *OALD9*'s entry "hashish" as an example, it is good to end with a warning ("Use of the drug is illegal in many countries", see Part A in Appendix I). However, its definition ("a drug made from the resin of the hemp plant, which gives a feeling of being relaxed when it is smoked or chewed") may embody a problematic positive attitude to drug use, and one could refine it by adding words like "misleading" or "dangerous" before "feeling". This is an implicit and subtle way of reframing.

The last two ways of reframing, questioning and neutralizing, concern controversial or sensitive issues. Questioning means asking a yes–no or rhetorical question. For instance, for the entry of "nature", one could use the question "Do you think man is good or evil by nature?" rather than a statement "She is evil by nature". Neutralizing refers to adopting a neutral stance when dealing with conflicting definitions by different communities and cultures. When dominant voices in society have dictated meanings of concepts/words at the expense of other social beliefs, it is advisable to listen to different voices with an ecological view. For instance, hunting is considered differently between wildlife conservatories (and governments) and local communities in Africa. The former only restrict it to the tracking and subsequent killing of game by "licensed" parties (usually foreigners), typically with rifles, all-terrain vehicles and professional trackers/rangers, and regard the same activity by members of local communities (usually with dogs, snares, spears and bows and arrows) as "poaching" (notwithstanding that the locals consider their own activities as hunting too). Besides giving a neutral definition (e.g. "go after wild animals to kill or catch them"), an African-oriented dictionary could point out the different understandings of the government and local communities to avoid biased representations of socio-cultural reality. The inclusion of conflicting definitions, as a way of dictionary customization in this case, may lead the government to rethink their policies, and at the same time enhance local communities' awareness of the divergence. Most importantly, the entry should give a warning against the brutal killing of animals at will, and clarify the differences in semantic prosody among hunting (neutral), poach (negative) and cull (positive). We do not mean that lexicographers should be preoccupied with pro-

viding only entries that are ecologically friendly. Sometimes, an eclectic and holistic approach is necessary for rebalancing cultural values for a sustainable society.[9]

5.3   Limitations and future research

Although the article offers refreshing insights into lexicographical research and practice, it is not free from limitations. The proposed approaches and models are still not substantiated with adequate empirical data from different types of dictionaries. The survey into "Big Five" dictionaries only covered 30 random pages from each of them. It was not an exhaustive retrieval of information for identifying all the destructive ideologies and frames. No investigation has been conducted into the intentions, attitudes and values of both dictionary users and designers. Furthermore, the five ways of reframing the dictionary discourse are far from enough to cover all the anti-ecological and anti-sociocultural problems.

There are theoretical and practical orientations for future research. Theoretically, eco-lexicographers still have to identify the principles similar to the succession and evolution of ecosystems, perhaps fruitful for illustrating the dynamism of different types of dictionaries. We need to reconsider, first, the values and concerns of traditional lexicography and, second, a context where ecologically oriented dictionaries compete with resources sustaining ideologies of consumption. Practically, more evidence should be collected to support the new paradigm of ecolexicography. Systematic investigations into dictionary discourse and dictionary use are needed by using techniques of data mining, machine learning and natural language processing. Different types or genres of dictionaries should be examined from an ecological perspective, and respective solutions can be found to improve them. A set of practical guidelines and methods for reframing the dictionary discourse should be developed.

Philosophically, we think that ecolexicography can gain inspiration from ancient Chinese worldviews of holism (focusing on the larger world than the body — the universe), interconnectedness, eclecticism and harmony (between the human and the cosmic, within society, and within the self). Although they might include "anthropocentric" interventions that entail ecologically constructive ideologies and practices, on the whole, they can help us engage in ecological awareness and deal with ecological crises. Different from the scientific tradition of viewing the world as separated physical parts and encouraging competitions (e.g. Darwinism), the Chinese cultural concept of "human-nature oneness" advocates altruism and tolerance (Lyu and Liu, in preparation). A harmonious view of language, mind and the world, and a new harmony of science, axiology and aesthetics are crucial (Huang and Zhao 2017; Zhou 2017) for a rebirth of lexicography in the digital era, an epoch of Anthropocene.

6.   Conclusion

To conclude, ecolexicography can be redefined as a new paradigm by adding symbolic and cognitive dimensions to the microlevel of a unified framework, and by upgrading the ontological, epistemological and methodological aspects related to this field. A redefined ecolexicography raises interesting questions. Besides proposing new terms including macro-ecolexicography and micro-ecolexicography, we have enriched the meaning of at least three groups of old theoretical terms or practices: (1) dictionary in/as a distributed cognitive system, distributed cognition, dictionary user identity, lexicographical interaction, dialogism and heteroglossia; (2) exoecological/endoecological position, values realizing, recontextualization; (3) frame analysis, user research, and lexical database construction. Ecolexicography as a novel paradigm is emancipatory, and could be a fruitful alternative to traditional practice and research, opening fresh paths and insights in an era of big data. It may help lexicographers solve the current problems with e-dictionaries in a new light, contributing to their role of serving the ecologies of language, socio-culture and nature. Additionally, it would be conducive to philosophical reflections on metalexicography.

Nevertheless, there is still a long way to go because there are many challenging issues. E-dictionary customization, for instance, is dependent on not only users' computer skills but also financial support for lexicographical projects. Dictionary compilation is limited primarily by the time and money available to do it.

Endnotes

1.   "Big Five" refers to the five best-known English learners' dictionaries: *OALD*, *LDOCE*, *COBUILD*, *CALD* (*Cambridge Advanced Learners' Dictionary*) and *MED* (*Macmillan English Dictionary*).

2.   As words are defined relative to frames, hearing/reading a word can activate its frame and the frames in its system in the brain (Lakoff 2010). Represented in syntactic patterns, frames involve semantic roles and their relations which are ultimately connected with people's cognitive frames (ibid.).

3.   With over 20 million users, *iCIBA* is said to be the second most popular dictionary app in China. Unfortunately, its latest version has no customization and interaction buttons as shown in Figure 4. Due to lack of data, we cannot find out if its designers would alter the entry in light of user feedback. We agree with one of the reviewers that designers' response to user feedback needs investigation because it enriches the meaning of dictionary editing/ revision as another potential basis for in-depth discussion under ecolexicography.

4.   One of the reviewers suggested considering "the possibility of enhanced methodologies incorporating advanced online reach at the data-gathering stage of compiling the dictionary" to make potential users participate in the creation (rather than revision) of the dictionary. We think that it is a promising area of research. Additionally, we agree with the reviewer that "not all users will have the magnanimity to participate constructively" — some would condemn the whole dictionary just because of one entry they dislike, and "instead of co-creating

meaning and value, they set out to defame the entire product and, thus, engage in 'destructive' tendencies against the dictionary". This proves the importance of ecolexicography in inculcating a sense of responsibility and correct values in dictionary users.

5. One reviewer holds that there may be ironic cases where an anti-ecological advertisement pops up ahead of ecologically sensitive entries, like dirty money sponsoring charity programs. This is one of the challenges eco-lexicography faces. Considering the varied nature and themes of advertisements, we suggest that dictionary developers be selective and refuse anti-ecological advertisements (see Dziemianko 2020 for the effect of advertising on online dictionary usefulness).

6. Customization of the interface style is important especially for the visually impaired or people with low vision who may wish to change the background color and light contrast (e.g. night theme) to adjust glare to read comfortably. We owe this idea to one of the reviewers.

7. Framing is the use of a story from one area of life (a frame) to structure how another area of life is conceptualized. Reframing is the act of framing a concept in a way that is different from its typical framing in a culture (Stibbe 2015: 47). A discourse can be reframed with concepts redefined for communicating new values.

8. One of the reviewers asked us to think of "he is evil by nature" as an illustrative example which may create a negative frame in the user's mind. One solution is to change the statement into a question (see Section 5.2).

9. We are thankful to one of the reviewers for the examples in this paragraph. S/he also mentioned the case of sanctioned culling of wildlife to reduce ballooning population sizes of specific species that threaten the environment. This kind of anthropocentricity, if unavoidable, may be justified. After all, humans form part of the ecosystem with other members of nature. Sanctioned culling of wildlife is different from animal abuse and killing at will. An ecological dictionary may allow for some flexibility and inclusiveness in treating entries that border on "ecologically (non)destructive ideologies". Eclecticism is a wise policy.

Acknowledgments

References

A.   Dictionaries

A New English–Chinese Dictionary (Century edition). 2003. *NECD* Compilation Group. Shanghai: Shanghai Translation Publishing House.

Buk Bang Sinda (*Bidayuh–Malay–English Dictionary*). 2013. Kuching, Malaysia: Dayak Bidayuh National Association.

Cambridge English Dictionary. Cambridge: Cambridge University Press. https://dictionary.cambridge.org/dictionary/english/. Accessed on 28 Jan. 2020.

Collins COBUILD Advanced Learner's Dictionary (8th edition, *COBUILD8*). 2014. Glasgow: HarperCollins.

Dictionary of Environment and Ecology (5th edition). 2004. Collin, Peter H. (Ed.). London: Bloomsbury.

Fayu Zhushou French Dictionaries. Shanghai Qianyan Network Technology Corporation. An app downloaded from https://www.francochinois.com/v4/fr/app/download#mobile. Accessed on 20 Jan. 2020.

Jinshan Ciba English Dictionaries (*iCIBA*). Jinshan Software Corporation. http://cp.iciba.com. Downloaded and consulted on 20 Jan. 2018.

Lexico.com (Oxford Dictionaries English). Dictionary.com and OUP. https://www.lexico.com/. Accessed on 12 Apr. 2020.

Longman Dictionary of Contemporary English. Pearson Education. http://www.ldoceonline.com/. Accessed on 25 Jan. 2020.

Longman Dictionary of Contemporary English (5th edition, *LDOCE5)*. 2009. Mayor, Michael and Chris Fox (Eds.). London: Pearson Education.

Merriam-Webster's Learner's Dictionary. http://www.learnersdictionary.com. Merriam-Webster, Incorporated. Accessed on 12 Apr. 2020.

Oxford Advanced Learner's Dictionary (9th edition, *OALD9*). 2015. Deuter, Margaret, Jennifer Bradbery and Joanna Turnbull (Eds.). Oxford: OUP.

Urban Dictionary. https://www.urbandictionary.com. Accessed on 12 Apr. 2020.

Wordsmyth Kids Dictionary (WILD). https://kids.wordsmyth.net/wild/. Accessed on 13 Feb. 2018.

B.   Other literature

Albuquerque, Davi B. 2018. As relações entre Ecolexicografia e Lexicografia Pedagógica (The Relations between Ecolexicography and Pedagogical Lexicography). *Domínios de Lingu@gem* 12(4): 2066-2101.

Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.

Bang, Jørgen Chr. and Wilhelm Trampe. 2014. Aspects of an Ecological Theory of Language. *Language Sciences* 41(A): 83-92.

Béjoint, Henri. 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.

Benson, Phil. 2002. *Ethnocentrism and the English Dictionary*. London: Routledge.

Bergenholtz, Henning and Sandro Nielsen. 2006. Subject-Field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.

Blackmore, Elena and Tim Holmes (Eds.). 2013. *Common Cause for Nature: Values and Frames in Conservation.* Machynlleth, Wales: Public Interest Research Centre.

Bosch, Sonja E. and Marissa Griesel. 2020. Exploring the Documentation and Preservation of African Indigenous Knowledge in a Digital Lexical Database. *Lexikos* 30: 1-28.

Chen, Wenge. 2015. Bilingual Lexicography as Recontextualization: A Case Study of Illustrative Examples in a New English–Chinese Dictionary. *Australian Journal of Linguistics* 35(4): 311-333.

Chen, Wenge. 2017. Lexicography, Discourse and Power: Uncovering Ideology in the Bilingualization of a Monolingual English Dictionary in China. *Pragmatics and Society* 8(4): 601-629.

Chen, Wenge. 2019. Towards a Discourse Approach to Critical Lexicography. *International Journal of Lexicography* 32(3): 362-388.

Cowley, Stephen J. 2011. Distributed Language. Cowley, Stephen J. (Ed.). 2011. *Distributed Language:* 1-14. Amsterdam/Philadelphia: John Benjamins.

Cronin, Michael. 2017. *Eco-Translation: Translation and Ecology in the Age of the Anthropocene.* Abingdon/New York: Routledge.

Dziemianko, Anna. 2020. Smart Advertising and Online Dictionary Usefulness. *International Journal of Lexicography* 33(4): 377-403.

Ezquerra, Manuel A. 1995. Political Considerations on Spanish Dictionaries. Kachru, Braj B. and Henry Kahane (Eds.). 1995. *Cultures, Ideologies and the Dictionary: Studies in Honor of Ladislav Zgusta:* 143-152. Tübingen: Max Niemeyer.

Fairclough, Norman. 1992. The Appropriacy of "Appropriateness". Fairclough, Norman (Ed.). 1992. *Critical Language Awareness:* 33-56. London: Longman.

Farina, Donna M.T.Cr., Marjeta Vrbinc and Alenka Vrbinc. 2019. Problems in Online Dictionary Use for Advanced Slovenian Learners of English. *International Journal of Lexicography* 32(4): 458-479.

Fellbaum, Christiane. 2014. Large-scale Lexicography in the Digital Age. *International Journal of Lexicography* 27(4): 378-395.

Fillmore, Charles J. 1985. Frames and the Semantics of Understanding. *Quaderni Di Semantica* 6(2): 222-254.

Fillmore, Charles J. and Collin F. Baker. 2009. A Frames Approach to Semantic Analyis. Heine, Bernd and Heiko Narrog (Eds.). 2009. *The Oxford Handbook of Linguistic Analysis:* 313-339. Oxford: Oxford University Press.

Fusari, Sabrina. 2018. Changing Representations of Animals in Canadian English (1920s–2010s). *Language & Ecology*: 1-32.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin.

Godwin-Jones, Robert. 2011. Emerging Technologies: Mobile Apps for Language Learning. *Language Learning & Technology* 15(2): 2-11.

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

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

Granger, Sylviane. 2012. Electronic Lexicography: From Challenge to Opportunity. Granger, Sylviane and Magali Paquot (Eds.). 2012. *Electronic Lexicography:* 1-11. Oxford: Oxford University Press.

Halliday, Michael A.K. 1985. *An Introduction to Functional Grammar.* London: Edward Arnold.

Hodges, Bert H. and Reuben M. Baron. 2007. On Making Social Psychology More Ecological and Ecological Psychology More Social. *Ecological Psychology* 19(2): 79-84.

Hoey, Michael. 2001. A Clause-Relational Analysis of Selected Dictionary Entries: Contrast and Compatibility in the Definitions of 'Man' and 'Woman'. Caldas-Coulthard, Carmen Rosa and Malcolm Coulthard (Eds.). 2001. *Texts and Practices: Readings in Critical Discourse Analysis:* 178-190. London: Routledge.

Hornscheidt, Antje. 2008. A Concrete Research Agenda for Critical Lexicographic Research within Critical Discourse Studies: An Investigation into Racism/Colonialism in Monolingual Danish, German, and Swedish Dictionaries. *Critical Discourse Studies* 5(2): 107-132.

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.

Huang, Fang and Sven Tarp. 2021. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvement. *Lexikos* 31: 68-92.

Huang, Guowen and Ruihua Zhao. 2017. On the Origin, Aims, Principles and Methodology of Eco-Discourse Analysis. *Modern Foreign Languages* 40(5): 585-596.

Jones, Mari C. (Ed.). 2015. *Policy and Planning for Endangered Languages*. Cambridge: Cambridge University Press.

Kravchenko, Alexander V. 2009. The Experiential Basis of Speech and Writing as Different Cognitive Domains. *Pragmatics & Cognition* 17(3): 527-548.

Kress, Gunther and Theo van Leeuwen. 2006. *Reading Images: The Grammar of Visual Design.* 2nd edition. London: Routledge.

Lakoff, George. 2010. Why It Matters How We Frame the Environment. *Environmental Communication* 4(1): 70-81.

Lakoff, George. 2014. *Don't Think of an Elephant! Know Your Values and Frame the Debate.* 2nd edition. Vermont: Chelsea Green Publishing.

Larsen-Freeman, Diane. 2011. Complex, Dynamic Systems: A New Trandisciplinary Theme for Applied Linguistics? *Language Teaching* 45(2): 202-214.

Lee, Sang Yong. 2017. *A Theoretical Model for a Dictionary of the Endangered Sherpa Language.* Unpublished M.A. thesis. Stellenbosch: Stellenbosch University.

Lew, Robert. 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.

Lew, Robert. 2012. How Can We Make Electronic Dictionaries More Effective? Granger, Sylviane and Magali Paquot (Eds.). 2012. *Electronic Lexicography:* 343-361. Oxford: Oxford University Press.

Li, Ping. 2012. Ideology-oriented Translations in China: A Reader-Response Study. *Perspectives* 20(2): 127-137.

Linell, Per. 2009. *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-Making.* Charlotte, NC: Information Age Publishing.

Linell, Per. 2013. Distributed Language Theory, with or without Dialogue. *Language Sciences* 40: 168-173.

Liu, Xiqin. 2015. Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos* 25: 210-232.

Liu, Xiqin. 2017. Multimodal Exemplification: The Expansion of Meaning in Electronic Dictionaries. *Lexikos* 27: 287-309.

Liu, Xiqin, Dongping Zheng and Yushuai Chen. 2019. Latent Classes of Smartphone Dictionary Users among Chinese EFL Learners: A Mixed-Method Inquiry into Motivation for Mobile Assisted Language Learning. *International Journal of Lexicography* 32(1): 68-91.

Lyu, Jing and Xiqin Liu. In preparation. For a Better World: Revealing Ecology-destructive Frames in English Learner's Dictionaries.

Makkai, Adam. 1993. *Ecolinguistics: Towards a New Paradigm for the Science of Language?* London/New York: Pinter Publishers.

Moon, Rosamund. 2014. Meanings, Ideologies and Learners' Dictionaries. Abel, Andrea, Chiara Vettori and Natascia Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15–19, 2014*: 85-105. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.

Ogilvie, Sarah. 2013. *Words of the World: A Global History of the* Oxford English Dictionary. Cambridge: Cambridge University Press.

Rossi, Fabio and Maria Grazia Sindoni. 2017. The Phantoms of the Opera: Toward a Multidimensional Interpretative Framework of Analysis. Sindoni, Maria Grazia, Janina Wildfeuer and Kay L. O'Halloran (Eds.). 2017. *Mapping Multimodal Performance Studies:* 61-84. New York/Abingdon: Routledge.

Sarmento, Manoel Soares. 2000. Ecolexicography: Words and Expressions We Should Live By. *Österreichische Linguistiktagung Graz, 8.–10. Dezember 2000. Akten des Symposiums 30 Jahre Sprache und Ökologie — Errungenschaften, Visionen.* Graz: Graz University.

Sarmento, Manoel Soares. 2002. Ecolexicography: Ecological and Unecological Words and Expressions. Fill, A., H. Penz and W. Trampe (Eds.). 2002. *Colourful Green Ideas: Papers from the Conference 30 Years of Language and Ecology (Graz, 2000) and the Symposium Sprache und Ökologie (Passau, 2001)*: 487-492. Viena: Peter Lang.

Sarmento, Manoel Soares. 2005. Por Uma Ecolexicografia. *Confluências* 2: 84-97.

Steffensen, Sune Vork. 2015. Distributed Language and Dialogism: Notes on Non-Locality, Sense-Making and Interactivity. *Language Sciences* 50: 105-119.

Steffensen, Sune Vork and Alwin Fill. 2014. Ecolinguistics: The State of the Art and Future Horizons. *Language Sciences* 41: 6-25.

Stibbe, Arran. 2014. Ecolinguistics and Erasure: Restoring the Natural World to Consciousness. Hart, Christopher and Piotr Cap (Eds.). 2014. *Contemporary Critical Discourse Studies:* 583-602. London: Bloomsbury Academic.

Stibbe, Arran. 2015. *Ecolinguistics: Language, Ecology and the Stories We Live By.* Abingdon/New York: Routledge.

Tarp, Sven. 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.

Tenorio, Encarnación Hidalgo. 2000. Gender, Sex and Stereotyping in *The Collins Cobuild English Language Dictionary*. *Australian Journal of Linguistics* 20(2): 211-230.

Tian, Yu, Xiao-wei Wang and Xin-qi Guo. 2016. On Ecological Views in Texts of Learner's Dictionaries and Their Educational Functions: A Text Analysis of *New Age English–Chinese Dictionary*. *Journal of Shanxi Datong University (Social Science Edition)* 30(6): 73-75, 79.

Trampe, Wilhelm. 2001. Language and Ecological Crisis: Extracts from a Dictionary of Industrial Agriculture (Translated by Peter Mühlhäusler). Fill, Alwin and Peter Mühlhäusler (Eds.). 2001. *The Ecolinguistics Reader: Language, Ecology and Environment:* 232-240. London/New York: Continuum.

Tsunoda, Tasaku. 2005. *Language Endangerment and Language Revitalization*. Trends and Linguistics Studies and Monographs 148. Berlin/New York: De Gruyter.

Van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London/New York: Routledge.

Van Lier, Leo. 2002. An Ecological-Semiotic Perspective on Language and Linguistics. Kramsch, Claire (Ed.). 2002. *Language Acquisition and Language Socialization: Ecological Perspectives:* 140-164. London/New York: Continuum.

Wang, Ning. 2003. On the Educational Function of Dictionaries: Comments on *New Century Modern Chinese Dictionary* by Tongyi Wang. *Lexicographical Studies* 1: 1-5.

Willinsky, John. 1994. *Empire of Words: The Reign of the OED*. Princeton, N.J.: Princeton University Press.

Winestock, Christopher and Young-kuk Jeong. 2014. An Analysis of the Smartphone Dictionary App Market. *Lexicography: Journal of ASIALEX* 1(1): 109-119.

Wurm, Stephen A. 2007. Threatened Languages in the Western Pacific Area from Taiwan to, and Including, Papua New Guinea. Brenzinger, Matthias (Ed.). 2007. *Language Diversity Endangered:* 374-390. Berlin/New York: Mouton de Gruyter.

Xu, Hai. 2009. *Towards Prototypical Exemplification in English Dictionaries for Chinese EFL Learners.* Beijing: Science Press.

Yong, Heming and Jing Peng. 2007. *Bilingual Lexicography from a Communicative Perspective.* Amsterdam/Philadelphia: John Benjamins.

Zhang, Yihua. 2015. *Second Language Acquisition and Learner's Dictionaries.* Beijing: The Commercial Press.

Zhang, Yiqiong, David Machin and Tao Song. 2015. Visual Forms of Address in Social Media Discourse: The Case of a Science Communication Website. *Journal of Multicultural Discourses* 10(2): 236-252.

Zheng, Dongping. 2012. Caring in the Dynamics of Design and Languaging: Exploring Second Language Learning in 3D Virtual Spaces. *Language Sciences* 34(5): 543-558.

Zheng, Dongping, Kristi Newgarden and Michael F. Young. 2012. Multimodal Analysis of Language Learning in *World of Warcraft* Play: Languaging as Values-Realizing. *ReCALL* 24(3): 339-360.

Zhou, Wenjuan. 2017. Ecolinguistics: Towards a New Harmony. *Language Sciences* 62: 124-138.

Appendix I: Destructive Instances in "Big Five" Dictionaries

The following definitions and examples are from 30 random pages in each of three English Learner's Dictionaries: *OALD9*, *LDOCE5* and *COBUILD8*. The coding scheme is as follows.

DictionaryXY: Z where Dictionary = *OALD9*, or *LDOCE5*, or *COBUILD8*; X = D (definition) or E (example); Y = sense number; Z = headword. For example, "OALD9E1: bam" means that it is an illustrative example from the first sense of the entry "bam" in *Oxford Advanced Learner's Dictionary*, 9th edition. For an entry with more than one instance, a number is given in square brackets to indicate the order. For instance, "OALD9E3[1]: abuse" means it is the first instance in this entry.

A.   *OALD9*

1.	OALD9E3[1]: abuse	All the children had been physically and emotionally abused.
2.	OALD9E3[2]: abuse	He had abused his own daughter (= had sex with her).
3.	OALD9E3[3]: abuse	The boy had been sexually abused.
4.	OALD9E2: accumulation	An accumulation of toxic chemicals.
5.	OALD9E1: bam	She pointed the gun at him and — bam!
6.	OALD9E3: bottle	After his wife died, he really hit the bottle (= started drinking heavily).
7.	OALD9E1: calling	He realized that his calling was to preach the gospel.
8.	OALD9E2: contrast	Her actions and her promises contrasted sharply.
9.	OALD9E2: desertion	She felt betrayed by her husband's desertion.
10.	OALD9E2: dirt	He picked up a handful of dirt and threw it at them.
11.	OALD9E3: dirt	Do you have any dirt on the new guy?
12.	OALD9E2[1]: environment	Pollution of the environment.
13.	OALD9E2[2]: environment	Damage to the environment.
14.	OALD9E6: escape	Toxic waste escaping into the sea.
15.	OALD9E8: fly	A large stone came flying in through the window.
16.	OALD9E1: give up on	His teachers seem to have given up on him.
17.	OALD9E5: graze	The bullet grazed his cheek.
18.	OALD9D1: hashish	A drug made from the resin of the hemp plant, which gives a feeling of being relaxed when it is smoked or chewed. Use of the drug is illegal in many countries.
19.	OALD9E1: helpless	It's natural to feel helpless against such abuse.
20.	OALD9E1: hurl	He hurled a brick through the window.
21.	OALD9E2: hurl	Rival fans hurled abuse at each other.
22.	OALD9E1: leakage	A leakage of toxic waste into the sea.
23.	OALD9E1: machine-gun	A group of prisoners was taken into the forest and machine-gunned.
24.	OALD9E4: mad	She's completely power-mad.
25.	OALD9E1: mar	The game was marred by the behaviour of drunken fans.

26.	OALD9E1: pit	The body had been dumped in a pit.
27.	OALD9E2: point	He pointed the gun at her head.
28.	OALD9E2: rake	They raked the streets with machine-gun fire.
29.	OALD9E1: raw	These fish are often eaten raw.
30.	OALD9D1: rubbish	Things that you throw away because you no longer want or need them.
31.	OALD9E1: sea	The waste was dumped in the sea.
32.	OALD9E1: treatment	Guests at the health spa receive a range of beauty treatments.
33.	OALD9E2: which	Houses which overlook the lake cost more.
34.	OALD9E4: wretched	Is it that wretched woman again?

B.   *LDOCE5*

35.	LDOCE5D2: amber	A hard yellowish brown substance used to make jewellery.
36.	LDOCE5D1: buckskin	Strong soft leather made from the skin of a deer or goat.
37.	LDOCE5E2: clutch	A small boy trying to escape from his mother's clutches.
38.	LDOCE5E2: coach	The child was coached for stardom by her mother.
39.	LDOCE5E1: compromising	The doctor was found in a compromising position with a nurse (= having sex with her).
40.	LDOCE5D2: crocodile	The skin of this animal, used for making things such as shoes.
41.	LDOCE5E2: crocodile	A crocodile briefcase.
42.	LDOCE5E2: discharge	The discharge of toxic waste into the sea.
43.	LDOCE5E3: experimentation	Experimentation with cannabis.
44.	LDOCE5E11: eye	We went to dances, but only under the watchful eye of our father.
45.	LDOCE5D1: gateway drug	A drug such as cannabis which is not a very dangerous drug, but which some people believe leads to the use of more dangerous drugs such as heroin.
46.	LDOCE5E7: go	He went crazy and tried to kill her.
47.	LDOCE5D1: hypnotic	A drug that helps you to sleep
48.	LDOCE5E1: hypocrisy	It would be sheer hypocrisy to pray for success, since I've never believed in God.
49.	LDOCE5E13: job	She looks completely different in this photo — she must have had a nose job.
50.	LDOCE5E1[1]: kill	Why did she kill her husband?
51.	LDOCE5E1[2]: kill	Murray held a gun to his head and threatened to kill him.
52.	LDOCE5D1: lanolin	An oil that is obtained from sheep's wool, and is used in skin creams.
53.	LDOCE5E4: lead	What led him to kill his wife?
54.	LDOCE5E1: motivate	We may never know what motivated him to kill his wife.
55.	LDOCE5E26: out	It's time we voted the Republicans out.
56.	LDOCE5D1: pigskin	Leather made from the skin of a pig.
57.	LDOCE5E1: pigskin	A pigskin suitcase.

58.	LDOCE5E3: practice	The practice of dumping waste into the sea.
59.	LDOCE5E2: raccoon	A raccoon coat.
60.	LDOCE5D1: sealskin	The skin or fur of some types of seal, used for making leather or clothes.
61.	LDOCE5D1: shammy	A piece of chamois leather, used for cleaning or polishing.
62.	LDOCE5E1: status symbol	A Rolls Royce is seen as a status symbol.
63.	LDOCE5E1[1]: staunch	A staunch conservative.
64.	LDOCE5E1[2]: staunch	One of Bush's staunchest supporters.
65.	LDOCE5E1: stay out	He started staying out late, drinking.
66.	LDOCE5E2: unreal	Many people go into marriage with unreal expectations.
67.	LDOCE5E3: well	I went out and got well and truly (= completely) drunk.
68.	LDOCE5E1: widespread	The widespread use of chemicals in agriculture.

C.   *COBUILD8*

69.	COBUILD8E1: admit	Up to two thirds of 14 to 16 year olds admit to buying drink illegally ...
70.	COBUILD8E6: aim	He was aiming the rifle at Wade.
71.	COBUILD8E1: air	... water and air pollutants.
72.	COBUILD8E1: arrestable	Possession of cannabis will no longer be an arrestable offence.
73.	COBUILD8E9: at	The crowds became violent and threw petrol bombs at the police ...
74.	COBUILD8E8: chase	... bear robes, mountain lion hides, and other trophies of the chase.
75.	COBUILD8E1: chop off	They dragged him to the village square and chopped his head off.
76.	COBUILD8E3: collect	After collecting the cash, the kidnapper made his escape down the disused railway line.
77.	COBUILD8E1: commence	The hunter knelt beside the animal carcass and commenced to skin it.
78.	COBUILD8E1: dissect	We dissected a frog in biology class.
79.	COBUILD8E2: fright	The snake picked up its head and stuck out its tongue which gave everyone a fright ...
80.	COBUILD8E1: gore	He was gored to death in front of his family.
81.	COBUILD8E2: gossipy	... gossipy old women.
82.	COBUILD8E3: grass	I started smoking grass when I was about sixteen. (grass = marijuana)
83.	COBUILD8E1: immanent	God is immanent in the world.
84.	COBUILD8E4: liberally	Chemical products were used liberally over agricultural land.
85.	COBUILD8E1: move in	Her husband had moved in with a younger woman ...
86.	COBUILD8E1: pizzazz	... a young woman with a lot of energy and pizzazz.
87.	COBUILD8E1[1]: prescribe	She took twice the prescribed dose of sleeping tablets ...

88.	COBUILD8E1[2]: prescribe	The law allows doctors to prescribe contraception to the under 16s.
89.	COBUILD8E1: repellent	... a very large, very repellent toad ...
90.	COBUILD8E5: sniff	He felt light-headed, as if he'd sniffed glue.
91.	COBUILD8E5: sniffer	... teenage glue sniffers.
92.	COBUILD8E1: struggle on	Why should I struggle on to please my parents? ...
93.	COBUILD8E2: salvation	... those whose marriages are beyond salvation.
94.	COBUILD8D1: sea	The sea is the salty water that covers about three-quarters of the earth's surface
95.	COBUILD8E5: trip	The biggest star perk, and the biggest power trip, must be the private plane.
96.	COBUILD8E7: trip	One night I was tripping on acid.
97.	COBUILD8E1: unclean	... the Western attitude to insects as being dirty and unclean ...
98.	COBUILD8E2: walk out	Her husband walked out on her ...
99.	COBUILD8E1[1]: weaponize	They were close to weaponizing ricin — a lethal plant toxin.
100.	COBUILD8E1[2]: weaponize	... the plan to weaponize outer space.
101.	COBUILD8E1: wearer	The mascara is suitable for contact lens wearers.
102.	COBUILD8E4: X	You can only make X amount of dollars a year.

Appendix II: Destructive Frames and Ideologies in "Big Five" Dictionaries

See Appendix I for the data source and the coding scheme.

A. Frames of Violence, Drugs and Alcohol: Violence and (Ab)use of Drugs and Alcohol Are Normal

  A-1. Frames of Violence: Violence Is Worth a Try; Some Groups of People Are Doomed to Be Hurt or to Be in Conflicts

(1) Sudden Attacks Are Normal Phenomena
- He picked up a handful of dirt and threw it at them. (OALD9E2: dirt)
- A large stone came flying in through the window. (OALD9E8: fly)
- The bullet grazed his cheek. (OALD9E5: graze)
- He hurled a brick through the window. (OALD9E1: hurl)

(2) Public Acts of Violence Are Normal Phenomena
- The crowds became violent and threw petrol bombs at the police ... (COBUILD8E9: at)

(3) People Can Be Killed at Will
- The body had been dumped in a pit. (OALD9E1: pit)
- They raked the streets with machine-gun fire. (OALD9E2: rake)
- A group of prisoners was taken into the forest and machine-gunned. (OALD9E1: machine-gun)
- They dragged him to the village square and chopped his head off. (COBUILD8E1: chop off)

(4) Using Guns Is Cool
- She pointed the gun at him and — bam! (OALD9E1: bam)
- He pointed the gun at her head. (OALD9E2: point)
- He was aiming the rifle at Wade. (COBUILD8E6: aim)
- Murray held a gun to his head and threatened to kill him. (LDOCE5E1[2]: kill)

(5) Married People Are Spouse Killers (Spouse Killing Frame)
- Why did she kill her husband? (LDOCE5E1[1]: kill)
- We may never know what motivated him to kill his wife. (LDOCE5E1: motivate)
- What led him to kill his wife? (LDOCE5E4: lead)

(6) Men and Women Kill Each Other (Gendercide Frame)
- She pointed the gun at him and — bam! (OALD9E1: bam)
- He pointed the gun at her head. (OALD9E2: point)
- He went crazy and tried to kill her. (LDOCE5E7: go)

(7) Children Are the Target of Sexual Harassment
- He had abused his own daughter (= had sex with her). (OALD9E3[2]: abuse)
- The boy had been sexually abused. (OALD9E3[3]: abuse)
- All the children had been physically and emotionally abused. (OALD9E3[1]: abuse)

(8) Due Punishment Can Be Avoided
- After collecting the cash, the kidnapper made his escape down the disused railway line. (COBUILD8E3: collect)
- It's natural to feel helpless against such abuse. (OALD9E1: helpless)

A-2. Frames of Drugs: Using Drugs Is a Good Experience; People Can Legally Be Drug (Ab)users

(1) Using Drugs Is a Good Experience
- a drug made from the resin of the hemp plant, which gives a feeling of being relaxed when it is smoked or chewed. (OALD9D1: hashish)
- a drug such as cannabis which is not a very dangerous drug, but which some people believe leads to the use of more dangerous drugs such as heroin (LDOCE5D1: gateway drug)
- One night I was tripping on acid. (COBUILD8E7: trip)
- He felt light-headed, as if he'd sniffed glue. (COBUILD8E5: sniff)

(2) People, Including Teenagers, Can Legally Be Drug (Ab)users
- ... teenage glue sniffers. (COBUILD8E5: sniffer)
- I started smoking grass when I was about sixteen. (grass = marijuana) (COBUILD8E3: grass)
- Possession of cannabis will no longer be an arrestable offence. (COBUILD8E1: arrestable)
- experimentation with cannabis (LDOCE5E3: experimentation)

A-3. Frames of Drinking: Drinking Heavily Is a Normal Behavior
- After his wife died, he really hit the bottle (= started drinking heavily). (OALD9E3: bottle)
- He started staying out late, drinking. (LDOCE5E1: stay out)
- I went out and got well and truly (= completely) drunk. (LDOCE5E3: well)
- Up to two thirds of 14 to 16 year olds admit to buying drink illegally ... (COBUILD8E1: admit)

B. Frames of Animals and Plants: Animals and Plants Are Resources for Human (Ab)use

B-1. Frames of Animals: Animals Are Resources for Human (Ab)use; Animals Are Bad

(1) Animals Can Be Eaten at Will
- These fish are often eaten raw. (OALD9E1: raw)

(2) Animals Can Be Killed at Will
- We dissected a frog in biology class. (COBUILD8E1: dissect)
- ... bear robes, mountain lion hides, and other trophies of the chase. (COBUILD8E8: chase)
- The hunter knelt beside the animal carcass and commenced to skin it. (COBUILD8E1: commence)

(3) Animals Are Raw Materials for Products
- leather made from the skin of a pig (LDOCE5D1: pigskin)
- a pigskin suitcase (LDOCE5E1: pigskin)
- the skin of this animal, used for making things such as shoes (LDOCE5D2: crocodile)
- a crocodile briefcase (LDOCE5E2: crocodile)
- a raccoon coat (LDOCE5E2: raccoon)
- the skin or fur of some types of seal, used for making leather or clothes (LDOCE5D1: sealskin)
- a piece of chamois leather, used for cleaning or polishing (LDOCE5D1: shammy)
- strong soft leather made from the skin of a deer or goat (LDOCE5D1: buckskin)
- an oil that is obtained from sheep's wool, and is used in skin creams (LDOCE5D1: lanolin)

(4) Animals Are Ugly, Dirty and Dangerous Things
- ... a very large, very repellent toad ... (COBUILD8E1: repellent)
- ... the Western attitude to insects as being dirty and unclean ... (COBUILD8E1: unclean)
- The snake picked up its head and stuck out its tongue which gave everyone a fright ... (COBUILD8E2: fright)
- He was gored to death in front of his family. (COBUILD8E1: gore)

B-2. Frames of Plants: Plants Are Resources for Human (Ab)use
- They were close to weaponizing ricin — a lethal plant toxin. (COBUILD8E1: weaponize)

The *New Online English–Georgian Maritime Dictionary* (NEGMD): Current State of the Project

Anna Tenieshvili, *Foreign Languages Department,*
Batumi State Maritime Academy, Batumi, Georgia
(a.tenieshvili@bsma.edu.ge) (anna_tenieshvili@yahoo.com)

Abstract: My practical training at the Maritime Transport Administration of Georgia in 2018 inspired the project of compiling the NEGMD. The project was boosted by an international grant of the European Lexicographic Infrastructure (ELEXIS), which led to an invitation to visit the Instituut voor Nederlandse Lexicologie in Leyden, the Netherlands.

The aim of this report on the compilation of the NEGMD is to show the state of this project from a practical point of view using concrete examples of terminological entries.

The project includes two main issues: the compilation of the dictionary itself and the coinage of new maritime terms in Georgian to fill the existing lexical/terminological gaps. It is of great importance for the field of maritime education and training in Georgia, for the whole maritime economy of the country and for the development of the Georgian language and, consequently, for the fields of Georgian linguistics and lexicography.

All issues related to this project including the criteria according to which it is being compiled and the information each terminological entry of the dictionary comprises, will be thoroughly covered. Perspectives on future dictionary development will be presented, illustrating it by concrete examples from the NEGMD.

Keywords: DICTIONARY COMPILATION, GUIDING PRINCIPLES, TERMINOLOGICAL ENTRIES, COINAGE OF GEORGIAN MARITIME TERMINOLOGY

Opsomming: Die *Nuwe Aanlyn Engels–Georgiese Maritieme Woordeboek (NEGMD): Stand van die projek.* My praktiese opleiding by die Maritieme Vervoeradministrasie van Georgië in 2018 het die projek om die NEGMD saam te stel, geïnspireer. Die projek is bevorder deur 'n internasionale toekenning van die Europese Leksikografiese Infrastruktuur (ELEXIS), wat gelei het tot 'n uitnodiging om die Instituut voor Nederlandse Lexicologie in Leiden, Nederland, te besoek.

Die doel van hierdie verslag oor die samestelling van die NEGMD is om die stand van die projek uit 'n praktiese oogpunt te toon met konkrete voorbeelde van terminologiese inskrywings.

Die projek omvat twee hoofkwessies: die samestelling van die woordeboek self en die skepping van nuwe seevaartterme in Georgies om die bestaande leksikale/terminologiese gapings te vul. Dit is van groot belang vir die vakgebiede van maritieme opvoeding en opleiding in Georgië, vir die algehele maritieme ekonomie van die land en vir die ontwikkeling van die Georgiese taal, en dus vir die vakgebiede Georgiese linguistiek en leksikografie.

Alle kwessies rakende hierdie projek, insluitende die kriteria waarvolgens dit saamgestel

word en die inligting wat elke terminologiese inskrywing van die woordeboek bevat, sal deeglik gedek word. Perspektiewe op toekomstige woordeboekontwikkeling sal aangebied word deur dit met konkrete voorbeelde uit die NEGMD te illustreer.

Sleutelwoorde: WOORDEBOEKSAMESTELLING, RIGLYNE, TERMINOLOGIESE INSKRY-WINGS, NUUTSKEPPINGS IN GEORGIESE MARITIEME TERMINOLOGIE

Introduction

The aim of the present report is to familiarize readers with the progress of work on the compilation of the NEGMD. The dictionary represents a pioneering project both for the fields of Georgian lexicography and Georgian maritime terminology. The reason for this is that previously, all maritime dictionaries that had been compiled in Georgia were translational dictionaries, giving one or several meanings of a particular term. The dictionary subject of this report has an explanatory character that is conditioned by the fact that term definitions are given both in English and Georgian, and also for the purpose of illustrating the terminology usage, the terms are provided with corpora examples. The information given in each terminological entry is organized in such a way that the student and general user are given an idea and understanding of the particular concept expressed by means of terms in the language. The compilation of this dictionary is guided by a terminological rather than a lexicographic approach. Since this is a specialized dictionary, the approach is therefore onomasiological, rather than semasiological, i.e. starting from the concept and moving to naming the concept by means of a term.

The project also addresses the coinage of missing Georgian maritime terminology with the purpose of filling existing lexical gaps, an issue that is especially relevant for terms of maritime navigation. During this process, I identify terms that do not have Georgian equivalents and through cooperation with the Georgian Linguistics Institute and the State Language Chamber of Georgia terms will be coined and added to the dictionary.

The NEGMD project

Several contributing factors led to the start of the NEGMD project. They can be formulated and explained in the following way:

One of the priorities of the Georgian economy and sustainable development is to develop the maritime field. The development of the maritime field is especially important for the economics of the country and this field should be developed in several directions. They are: ports, logistics, maritime education and training that comprises navigation, maritime transportation, maritime management and marine engineering. The importance of the maritime field is confirmed by the fact that this field makes a large contribution to the country's

economy and budget and therefore its development is of strategic importance.

Nowadays, the maritime field is an international field in which various countries of the world fulfil various functions. Georgia mainly fulfils a transit function and owing to its geostrategic location throughout history, the country has formed a link between Europe and Asia. Georgia, being an important centre of maritime education, can increase its role in the preparation of highly qualified seafarers since maritime education plays a significant role in establishing Georgia in the international domain and in the development of the country's economy.

The idea of the project of compiling the NEGMD occurred to me after I had my practical training at the Maritime Transport Administration of Georgia in 2018. This practical training was obligatory since, in addition to my PhD in English Philology, I received an MSc in Maritime Affairs majoring in Maritime Education and Training from the World Maritime University (WMU).

As part of the above-mentioned project I received an international grant from the European Lexicographic Infrastructure (ELEXIS) in 2019 and being the receiver of the grant program, I was invited to the Instituut voor Nederlandse Lexicologie in Leyden, the Netherlands. In order to compile a dictionary that will contain various types of information including definitions, corpus examples, some encyclopedic information, related words, etc., it was necessary to conduct some research work, to pay attention to advice of specialists experienced in similar work. As my dictionary is intended to have both terminological and lexicographic features, the opinions of terminologists and lexicographers are to be taken into consideration when composing entries and compiling the dictionary as a whole. In addition to the dictionary compilation, the purpose of my work is to conduct research in order to fill in the lexical gaps existing in maritime terminology in Georgian.

During work on the NEGMD project, I am guided by the following recommendations for the compilation of specialized dictionaries that were given by L'Homme (2006: 182) in the article "The Processing of Terms in Dictionaries":

— Dictionaries should consider user needs and include highly specialized but also less specialized items;
— Terminologists or specialized lexicographers should make more use of evidence found in corpora as a basis for taking decisions about terms;
— Dictionaries should include more data on terms (e.g., collocations, valence patterns, images);
— Bilingual and multilingual dictionaries should account for interlinguistic differences;
— Dictionaries should describe relationships between terms;
— Specialized dictionaries should contain encyclopedic or pragmatic information;
— Definitions should be structured in order to display key conceptual components.

The project includes consideration of and work on maritime terminology. As maritime terminology is divided into two parts, general maritime terminology and specific maritime terminology, I started the compilation of the dictionary from work on general maritime terminology. At the present stage, the project comprises more than 1000 terminological entries, the majority of which are general maritime terms.

The following criteria and principles guide the compilation of the dictionary:

1. The terminology comprises the following maritime subfields: maritime navigation, marine mechanical engineering, marine electrical engineering, logistics, economics of maritime transport, mechanization of port operations, cargo work and maritime law;
2. At the current stage I consider and work on simple terms, as for terminological lemmas, they will form the following stage of the dictionary compilation work;
3. Each terminological entry includes definitions and corpus examples;
4. Each term is considered from the viewpoint of its monosemantic nature.

During dictionary compilation, I am guided by the international standard of terminology processing and enter all information related to terminological entries into a specially developed online platform.

As mentioned in Tenieshvili (2020: 485-491) the NEGMD project is very important to Georgia as a sovereign country, for Georgian linguistics and lexicography and for the Georgian maritime semantic field. Georgia is a sovereign country now, and terminology is one of the means of establishing the country as an independent state. Although maritime terminology is the most globalized semantic field in the world, each country, especially maritime nations like Georgia, must have its own maritime terminology. It is especially important for reaching the following aims:

1. Stimulation for developing Georgian maritime terminology;
2. Application of the developed maritime terminology in textbooks, maritime documentation and materials;
3. Economy of space and facilitation of understanding among specialists and students of the maritime field;
4. Improvement of the system of maritime education and consequently of the national maritime industry;
5. Contribution to the lexicography of Georgian;
6. Development of the lexicography of the maritime semantic field;
7. Establishing a basis for a Georgian maritime corpus with further integration of the Georgian maritime corpus into the general corpus of the Georgian language;
8. General contribution to Georgian linguistics.

It would be expedient to give an idea of the NEGMD project by demonstrating terminological entries that have already been developed in the present report:

accommodation n
/əˌkɒm.əˈdeɪ.ʃən/
—

საცხოვრებელი გემზე
fitted with cabins and catering facilities for offshore crews

საცხოვრებელი სათავსი გემზე ეკიპაჟის წევრებისთვის ან მგზავრებისთვის

Corpus examples
Accommodation was clean but basic.
გემზე მეზღვაურებისთვის ხელმისაწვდომია შესაბამისი საცხოვრებელი ფართობი.

afterpeak n
/af-ter-peek/
—

ახტერპიკი, ახტერშტევენი
the extreme after part of the interior of a hull, especially that part below the water immediately forward of the sternpost (opposed to forepeak)

სამოქალაქო გემის კიჩოს ტრიუმის განაპირა ნაკვეთეური; წარმოადგენს ბალასტის ცისტერნას და განკუთვნილია გემის გასაწონასწორებლად, დიფერენტის თავიდან ასაცილებლად

Corpus examples
Forepeak tanks are at the fore end of the hull and afterpeak tanks are at the after end.
ახტერშტევენი ერთხრახნიანი გემის არაბალანსირებული საჭით წარმოადგენს ორტოტიან ჩარჩოს.

anchorage n
[ˈæŋkərɪdʒ]
—

ლუხასადგომი
place where a boat is or can be anchored

ადგილი სადაც გემს შეუძლია ლუხის ჩაშვება ანუ გემების სადგომი ადგილი

Corpus examples
We are in between the Hospital and the Alaska Regional Medical Plaza. If you are approaching anchorage from the NORTH on the Glen Highway, turn left at the 2nd set of lights onto Airport Heights Road.
საკონსულო თანამდებობის პირს უფლება აქვს საჭირო დახმარება გაუწიოს და ხელი შეუწყოს წარმდგენი სახელმწიფოს გემს, რომელიც იმყოფება ადგილისამყოფელი სახელმწიფოს ნავსადგურში ან ლუხასადგომზე ან ადგილისამყოფელ სახელმწიფოს შიდა და ტერიტორიულ წყლებში.

charter n
/ˈtʃɑː.tər/

—

ფრახტი, გემის დაფრახტვა
renting of a plane or ship

გემით ტვირთის გადაზიდვის ქირა, საფასური

Corpus examples
Masters Championship in Cyprus: If you wish to book a charter, please contact ILCA as soon as possible. We expect all charter boats to have been booked by mid-January 2010, so it is advisable to contact ILCA now to avoid disappointment.
ერთ ნავს შეუძლია ზიდოს 100 ფუთიანი ტვირთი და იმისდამიხედვით, თუ რა რაოდენობისაა იგი ფრახტი არტვინიდან ბათუმამდე ღირს 10-დან 15 მანეთამდე.

[Charter Party](#)
[to charter](#)
[chartering](#)
col v n

exhaust n
/ɪɡˈzɔːst/

—

გამონაბოლქვი, გამობოლქვა
escape of used gas or vapor from an engine; gas or vapor thus escaping

ძრავას გამონაბოლქვი, გამონაბოლქვი/ნამუშევარი აირი ან ორთქლი

Corpus examples
As the bell is *very much larger* than the pipe feeding it with the hot exhaust, the hot exhaust does not rush through it as might be first assumed.
ეკოლოგიისა და მოსახერხებლობის თვალსაზრისით საუკეთესოდ ითვლება. ამ სიმაღლეზე ვერ აღწევს მანქანების გამონაბოლქვი, ხოლო ქარხნული და სხვა კვამლის კონცენტრაცია მაღალ სართულებზე მეტია.

helm n
[helm]

—

შტურვალი
handle or tiller, in large ships the wheel, by which the rudder is managed, sometimes extended so as to include the whole steering gear

გემის კურსზე მოძრაობის მართვის ორგანო, სახელურებიანი (ან უსახელურებო) ბორბალი, რომელიც სხვადასხვა კონსტრუქციის ამძრავით შეერთებულია საჭესთან

Corpus examples
Very few helms were sailing their boats anywhere near to the limit of their craft's capability, including some old hands.
ცალი ხელით თეთრად შეღებილი რკინის საწოლის თავგისოსი მტკიციდ ეპყრა, როგორც შკიპერს შტორმში შტურვალი ატორტმანებული ხომალდისა.

(extracted from: The NEGMD)

Along with the compilation of the dictionary itself, there is a second issue: the coinage of Georgian maritime terminology. There are different methods of terminology creation. Sometimes even calques can be used. In my opinion, as English is the official language of the maritime field established by the International Maritime Organization (IMO) based in London, it would be expedient if calquing takes place from the English language. That would contribute to the internationalization of terminology and this, in its turn, would facilitate communication between specialists of the maritime field. It is also important to ensure the coexistence of national terms and terms that have recently entered the language.

I think that when creating Georgian maritime terminology, specialists should be guided by the rules of Georgian and it would be better to avoid calquing taking into consideration international linguistic rules at the same time. Such a combined approach would help in reaching a certain balance in this issue. As an example, I would like to mention that owing to linguistic specifics of the English language there exist terms based on the noun + noun model in it, yet it does not mean that grammatical calques should be borrowed in line with lexical calques too.

In my opinion, the issue of coinage of the Georgian maritime terminology implies several organizational issues, such as:

1.   Revision of existing maritime dictionaries and glossaries;
2.   Selection of a policy for the coinage of maritime terms;
3.   Selection of methods for the coinage of maritime terms;
4.   Organization of a team.

Perfection and adoption of maritime terminology will contribute to the development of the entire maritime field and also its different subfields. It is very important to ensure the improvement of the academic level of maritime specialists.

Conclusion

Compilation of the *New Online English–Georgian Maritime Dictionary* is of utmost importance for the field of maritime education and training and for the whole maritime field of Georgia. It will stimulate the educational process in Georgian maritime educational institutions, provide much new information and increase comprehension of maritime issues, contributing to a better com-

prehension of maritime phenomena by Georgian maritime students in their native language. In addition, the project and all issues discussed in this report will contribute to the fields of lexicography and linguistics in Georgia, as it will enable native Georgians to study and comprehend maritime phenomena in their native language via correspondent terms and not only on the basis of English–English explanations, the practice that exists today.

The issues of compilation of the maritime dictionary and refining of Georgian maritime terminology are very important for the development of maritime education and for the establishment of Georgia in the international maritime arena.

References

L'Homme, Marie-Claude. 2006. The Processing of Terms in Dictionaries: New Models and Techniques. A State of the Art. *Terminology* 12(2): 181-188

New Online English–Georgian Maritime Dictionary. Available at: https://www.lexonomy.eu/zy44hrpwh/.

Tenieshvili, A. 2020. The New Online English–Georgian Maritime Dictionary Project. Challenges and Perspectives. Gavriilidou, Z., M. Mitsiaki and A. Fliatouras (Eds.). 2020. *Proceedings of the XIX Euralex International Congress, Alexandroupolis, Greece, 7–11 September 2021. Volume 1*: 485-491. Komotini, Greece: Democritus University of Thrace. Available at: https://euralex2020.gr/wp-content/uploads/2020/11/EURALEX2020_ProceedingsBook-p485-491.pdf.

Lexicographic Data Boxes
Part 1.
Lexicographic Data Boxes as
Text Constituents in Dictionaries[*]

Rufus H. Gouws, *Department of Afrikaans and Dutch,*
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)
and
D.J. Prinsloo, *Department of African Languages, University of Pretoria,*
Pretoria, South Africa (danie.prinsloo@up.ac.za)

And they all get put in boxes, little boxes all the same
(Malvina Reynolds)

Abstract: This article, the first in a series of three on lexicographic data boxes, focuses primarily on the occurrence of lexicographic data boxes as text constituents in dictionaries. Following a brief look at what data boxes are, the focus shifts to the different venues where these boxes can be accommodated within the central list of a dictionary. Boxes containing items and/or item texts can be positioned within articles, or article-externally as phased-in inner texts within a partial article stretch of a dictionary. Data boxes are used to convey data that need to be highlighted and are therefore often formally marked (a coloured background or within a frame) and are put in an article slot that has a position of salience. As dictionary entries they can participate in procedures of both lemmatic and non-lemmatic addressing. It is shown that a box should preferably be inserted close to its address. In articles of polysemous words, the user should unambiguously know for which sense(s) the box is relevant. As phased-in inner texts data boxes can be addressed at a lemma within the same partial article stretch but also, in the case of synopsis boxes, at lemmata in other article stretches. This demands procedures of remote addressing.

Keywords: ADDRESSING, ARTICLE STRETCH, ARTICLE WINDOWS, ARTICLE-EXTERNAL DATA BOXES, ARTICLE-INTERNAL DATA BOXES, DATA BOXES, DATA DISTRIBUTION, EXPANDED WORD LIST, INSERTS, LEXICOGRAPHIC DATA BOX, PARALLEL MACROSTRUC-TURE, PARTIAL ARTICLE STRETCH, PHASED-IN INNER TEXTS

Opsomming: Leksikografiese datakassies. Deel 1. Leksikografiese datakassies as tekskonstituente in woordeboeke. Hierdie artikel, die eerste in 'n reeks van drie oor leksikografiese datakassies, fokus veral op die voorkoms van leksikografiese datakassies as

---

[*]   This is the first in a series of three articles dealing with various aspects of lexicographic data boxes.

tekskonstituente in woordeboeke. Na 'n kort bespreking van wat datakassies is, verskuif die fokus na die verskillende plekke waar hierdie kassies in die sentrale woordelys van 'n woordeboek geakkommodeer kan word. Datakassies wat aanduiders en/of aanduidertekste bevat, kan binne-in artikels, of artikel-ekstern as ingefaseerde bbinnetekste in die deeltrajekte van 'n woordeboek geplaas word. Datakassies word gebruik om data oor te dra wat beklemtoon moet word en word daarom dikwels gemerk (met 'n gekleurde agtergrond of in 'n raam) en word in 'n artikelgleuf geplaas wat in 'n posisie is wat die aandag trek. As woordeboekinskrywings word datakassies betrek by prosedures van lemmatiese en nielemmatiese adressering. Dit word aangetoon dat 'n kassie liefs so na as moontlik aan sy adres geplaas moet word. In artikels van polisemiese woorde moet die gebruiker ondubbelsinnig kan weet vir watter betekenisonderskeiding(e) die kassie relevant is. As ingefaseerde binnetekste kan datakassies aan 'n lemma in dieselfde deeltrajek geadresseer wees, maar ook, in die geval van sinoptiese kassies, aan lemmata in ander artikeltrajekte. Dit vereis prosedures van verwyderde adressering.

Sleutelwoorde: ADRESSERING, ARTIKELDEELTRAJEK, ARTIKEL-EKSTERNE DATAKASSIES, ARTIKEL-INTERNE DATAKASSIES, ARTIKELTRAJEK, ARTIKELVENSTERS, DATAKASSIES, DATAVERSPREIDING, INGEFASEERDE BINNETEKSTE, INVOEGING, LEKSIKOGRAFIESE DATAKASSIE, PARALLELLE MAKROSTRUKTUUR, UITGEBREIDE WOORDELYS

1.   Introduction

The emergence[1] and establishment of dictionary structures as a focal area of research in the field of metalexicography and dictionary research has had a significant influence on the theory and practice of lexicography. The scope of this influence included a critical analysis and discussion of the occurrence and positioning of a variety of text compound constituents, texts and textual segments in dictionaries and dictionary articles. Research into dictionary structures had both a contemplative and a transformative approach, cf. Tarp (2004: 224). Existing structures were studied and described but a whole range of new structures were identified, proposed and employed in the planning and compilation of dictionaries. Research into dictionary structures resulted in lexicographers realising all articles in a given dictionary do not necessarily have to display exactly the same structure. A sequence of articles in a single article stretch does not merely have to be a presentation of more of the same. Instead of a single and consistently applied homogeneous article structure the lexicographer has the liberty to opt for less rigid heterogeneous article structures in which the obligatory microstructure, the default microstructure of the specific dictionary, can be supplemented by items enabling an extended obligatory microstructure. Different types of word list, micro- and article structures give lexicographers the opportunity to present partial article stretches, microstructural items and item texts in a way that would best fit the needs and reference skills of their intended target users. Yet again, the user-perspective plays a dominant role in the planning and compilation of dictionaries. This perspective also co-determines the data distribution and data presentation in any given

dictionary as a search region, the word list as a search field, an article as a search area and the article slots as search zones, but also the structure allocated to any article as well as the potential inclusion of phased-in inner texts in a given article stretch. For the study of dictionary structures the occurrence and positioning of each partial article stretch, item text, item and indicator is of significance. This also applies to data boxes as lexicographic text constituents.

Lexicographic data boxes are lexicographic text constituents frequently employed in the presentation of data in various search venues of dictionaries. Although data boxes have become a common phenomenon in dictionaries, relatively little attention has been paid to their presentation, the motivation for their use and the type of data a dictionary displays in this specific way, cf. Gouws and Prinsloo (2010) and Taljard, Prinsloo and Gouws (2014). This article firstly motivates the need for a theoretical discussion of data boxes, focusing on the types of data boxes and the textual positions allocated to them before following a contemplative approach by taking a look at some aspects of the current use of data boxes. A transformative approach is also followed and some ideas deemed necessary for an improved use of this lexicographic text segment will be presented in the final article of this three-part series.

The focus in this article is on the occurrence of data boxes as text constituents in dictionaries. In the follow-up articles, part 2 and 3 in this series, the focus will be on the contents of data in boxes and a look at future improvements respectively.

2.   What are lexicographic data boxes?

Before proceeding with a discussion of lexicographic data boxes it is important to have a clear understanding of what is meant in this article by the term *lexicographic data box* and its shortened form *data box*.

A lexicographic data box, or just data box, is a data-carrying constituent of the word list of a dictionary. It contains data presented

— as part of the treatment of a specific lemma that is the guiding element of the article in which the data box appears or that is in close proximity to the data box;
— as part of the treatment of various lemmata where the data box is an entry in an article or in a partial article stretch that accommodates one of the lemmata for which the treatment in the data box is relevant;
— in a phased-in inner text in an article stretch that contains a lemma for which the contents of the data box is relevant.

Data boxes contribute to the lexicographic treatment of a lemma in the same article stretch where they appear, specifically a lemma that is the guiding element of the article in which the data box is presented as entry or a lemma in close proximity of that data box. In addition, synoptic data boxes can also contain data relevant to the lexicographic treatment of lemmata in other article

stretches of the dictionary. Data boxes are usually distinguished from other data-carrying constituents in terms of the form in which they are presented. They often occur within a framed box, as a text constituent with a coloured background, as an article window (cf. Wiegand and Gouws 2011: 281), or in clearly identified phased-in inner texts that split a partial article stretch. Although data boxes often contain data presented within a frame or against a coloured background, this is not always the case. Some data boxes are not presented in boxed or coloured format.

The term *data box* is preferred to the frequently used term *text box* because texts are not the only type of data to be accommodated in these boxes.

Examples of different types of data boxes will be presented in later sections of this article.

3.   Why data boxes

Lexicographers should refrain from procedures that could result in a situation of data overload in their dictionaries, cf. Gouws and Tarp (2017). In its default microstructural presentation a dictionary article should contain the data the lexicographer regards as sufficient to satisfy the lexicographic needs of the envisaged target user. In addition to the data presented in the word list of a dictionary the lexicographer may also use a data distribution structure that allocates certain data to texts in the front and/or back matter sections of the dictionary. Where a lexicographer is convinced that specific data entries are needed to improve the lexicographic presentation and treatment in the word list of a given dictionary but the default microstructure, article structure and word list do not offer an appropriate position for such data entries, the lexicographer may embark on using data boxes — either article-internally or as phased-in inner texts. The use of data boxes needs to ensure an added value to the default treatment on offer in a given dictionary and the contents of data boxes should not be more of the same with regard to the default data presentation in dictionaries.

Data boxes are salient dictionary entries and as such they should be employed when there is a need to bring a non-default type of data to the attention of the user or to place more than the usual focus on a specific data item. As a result of lexicographic procedures used in an extended obligatory microstructure, care should be taken that data boxes do not become part of the obligatory microstructure and in so doing lose their significance and decrease the emphasis on the data included in these boxes. Lexicographers should make a clear distinction between using data boxes and using other lexicographic procedures, e.g. a system of labelling, to focus the attention of the user on a specific item.

4.   The need for a discussion of data boxes

The broad field of lexicography is characterised by an interactive relation

between theory and practice. Because the theory of lexicography emerged much later than the practice there had to be a lot of catching up to ensure a comprehensive theoretical coverage of the endeavours of the lexicographic practice and to present future compilers of dictionaries with the necessary theoretical basis for their dictionaries. In lexicography theory and practice do not develop in a parallel way. In lexicographic theory new suggestions come to the fore and they are often only applied in practice at a much later stage, e.g. the proposal for semi-integrated microstructures (Wiegand 1996). Some suggestions resulting from research in the field of theoretical lexicography never even find an application in the lexicographic practice. Practical lexicographers often introduce innovative approaches in their dictionaries and the theory may eventually include a discussion of these approaches when they have already been firmly established in dictionary compilation processes. Yet, it remains important that there should be a relation of reciprocity between theory and practice in lexicography: theory should learn from and influence the practice and the practice should learn from and influence the theory.

The occurrence of data boxes in printed dictionaries, especially learner's dictionaries, as well as in electronic dictionaries, especially those which are based on paper dictionaries, has increased significantly in the last decades. Although (learner's) dictionaries continuously witness an enhancement in quality, also as a result of extensive research in the field of theoretical lexicography with regard to this dictionary type, the competitive market and the competition between publishing houses have also had an influence on changes in the structures and contents of dictionaries. Many of these changes have added value to the dictionaries but in some instances they were, according to Wiegand and Gouws (2011: 238), not much more than trends in lexicographic face-lifting. The lexicographic practice did, however, introduce innovative uses of data boxes coming to the fore in many dictionaries. In this regard practice took the lead and theory unfortunately failed to respond quickly enough. Consequently many of these approaches have not yet been sufficiently appreciated and discussed by metalexicographers.

In order for lexicographic theory to keep up with developments in the practice and in order to ensure that future dictionary compilers can use the theory for both a contemplative and a transformative commitment to the planning and compilation of their dictionaries, the use of data boxes needs to be included in theory-based discussions.

5.   Data boxes as dictionary entries

5.1   Data boxes and addressing relations

The addressing structure and various addressing relations in dictionaries are important to identify the scope of each item in a dictionary article and the target of its treatment. According to Hausmann and Wiegand (1989: 328) a treat-

ment unit results "when a form mentioned and information relating to that form are brought together. The relation of form and information is that of topic and comment." The "way in which a form and information relating to that form are brought together is the *addressing procedure*." Each information item is addressed to a form that is known as its address.

Data boxes are functional constituents of dictionaries and should not be employed as mere procedures of lexicographic face-lifting. They are included in dictionaries as part of the lexicographic treatment procedures. When allocating a data box to a specific search zone cognizance needs to be taken of the specific treatment contributed by the data box. Dictionary users should know exactly what constitutes both the address and the addressing relation of each data box.

To ensure an optimal comprehension of the relevant addressing relation, it is important that the notion of addressing has to be employed in an unambiguous way. In this article the procedures of addressing as discussed in Gouws (2014) will be employed. Wiegand and Gouws (2013) restrict addressing to relations within condensed articles where items display the addressing relations. Wiegand and Gouws (2013: 273) say that there are no addressing relations in non-condensed dictionary articles or in other non-condensed accessible entries. This implies that an item like a lemma sign cannot be addressed by an item text, even if the item text contributes to the treatment of the lemma. In contrast to this approach, Gouws (2014: 183) expands the application of the notion of addressing. The term addressing is used to incorporate both primary addressing which is the traditional procedure prevailing in condensed articles, and secondary addressing to go beyond textual condensation and item status as prerequisites for addressed entries. In accordance with this approach both items and item texts can participate in procedures of addressing. The addressing element and its address do not have to be in the same dictionary article and addressing can occur between different types of constituents of a word list, e.g. also between a phased-in inner text and an item in a dictionary article.

5.2   The position for data boxes and the different types of data box entries

In accordance with the data distribution structure of a specific dictionary the lexicographer has to allocate data-carrying segments to the different search positions in the dictionary. Where a dictionary displays a frame structure the positioning of data in the articles of the word list will be complemented by text constituents or other entries presented in the front and/or back matter sections of that dictionary. As a carrier of text types a dictionary could also include outer texts in its middle matter section.

Dictionaries have different search positions, traditionally ranging from a search zone within a dictionary article to the article as search area and the word list, the largest search position, as search field, cf. Wiegand, Beer and Gouws (2013: 63). Gouws (2018: 228) argued in favour of a further and more compre-

hensive search position, i.e. the search region. This search position is consti-
tuted by all the textual components of a dictionary as a text compound. For a
single volume dictionary this is the most comprehensive search position.
Although the outer texts of the front and back matter sections of a dictionary
fall outside the domain of the search field they are part of the search region of a
dictionary. Data boxes typically occur within a word list, that is a search field,
of a dictionary. This could be within the central and only word list which is the
most frequent occurrence of data boxes, or within any other word list that is
part of a word list series of a given dictionary. Looking at data boxes in this
article, the focus will only be on those data boxes occurring in a word list, i.e.
within the search field of dictionaries. Data boxes in outer texts will not be dis-
cussed.

As entries in a search field the status of all data boxes is not the same. Data
boxes could be phased-in inner texts, article-internal item texts or mere items
occupying a search zone in a dictionary article. These different types of data
boxes will be discussed in the subsequent sections.

5.3   Expanded word lists

The data distribution structure of any dictionary determines the way in which
the lemmata presenting the macrostructural coverage of the dictionary have to
be ordered within the word list of that dictionary. Whether it is a straight
alphabetical ordering or an ordering with sinuous lemma files that enables the
use of niched and nested lemmata, a word list will consist of a number of article
stretches that accommodate articles with the lemmata as their guiding elements.

When planning the data distribution structure of a dictionary lexicographers
need to make a decision regarding the type of word list the dictionary will dis-
play. This is determined by specific features of the nature of data allocated to
the word list. In this regard a distinction is made between a single word list
and an expanded word list, cf. Bergenholtz, Tarp and Wiegand (1999: 1766). A
single word list contains article stretches but no inserts or phased-in inner texts,
whereas an expanded word list also contains article stretches as well as inserts
and/or phased-in inner texts. These inserts and phased-in inner texts split
sequences of articles within partial article stretches. Therefore an expanded
word list is also known as a split word list.

The distinction between inserts and phased-in inner texts is significant for
the identification of data boxes.

5.3.1 Distinguishing between inserts and phased-in inner texts

5.3.1.1 Inserts

According to the *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary*

of Lexicography and Dictionary Research (Wiegand et al. 2017), (henceforth abbreviated as WLWF) an insert is a text or text part that is inserted into the word list of a dictionary. It is an immediate constituent of the dictionary as a text compound. Inserts often are sections of photos, inserted between two article stretches or between two pages of an article stretch. In the *Merriam-Webster's Advanced Learner's English Dictionary* (Perrault 2008), (henceforth abbreviated as MWALD**),** the article stretch of the letter M is split between the lemmata *mascot* and *masculine* by an insert titled *Color Art.* This insert, with its own table of contents, contains pictures of themes like colours, vegetables, landscapes, gems and jewellery and clothing. These pictures do not adhere to the alphabetical ordering within the article stretch and although it splits the word list this insert is not an immediate constituent of either the specific article stretch or the word list. It is an immediate constituent of the dictionary as text compound and carrier of text types. Consequently inserts cannot be regarded as a type of data box or the type of data box discussed in this article.

The form of inserts sometimes resemble that of data boxes, e.g. the section on birds in an insert from the MWALD contains pictures of a variety of birds and the data box occurring in the article of the lemma *deer,* cf. figure 18 in paragraph 5.5.2.1, contains a number of pictures of different types of deer. Although both these text constituents are carriers of pictures, the insert is a different type of lexicographic text and will not be discussed any further in this article.

5.3.1.2 Phased-in inner texts

Phased-in inner texts occur in expanded word lists where they split the article stretches, resulting in internally expanded article stretches (Bergenholtz, Tarp and Wiegand 1999: 1766). The phased-in inner texts typically contain data relevant to a lemma in close proximity within the same article stretch and often also to other lemmata occurring in remote articles. The data in the phased-in inner texts can also have a more general relevance to the specific lemma or might link thematically with it. Phased-in inner texts often have an inner text title that can concur, but not necessarily, with the lemma of an article in its proximity. This title could also identify the more general nature of the data in the specific text. These texts are usually typographically distinguishable from default articles as constituents of a word list by being framed or presented with a coloured background. Examples will be provided in a subsequent section of this article.

Article-internal data boxes and phased-in inner texts sometimes show strong resemblances but the nature of a specific box as a text constituent of the word list determines whether it has to be regarded as an article-internal box or a box presented as phased-in inner text.

5.4   Article-internal data boxes

5.4.1   Items and item texts

One of the most frequent occurrences of data boxes is within dictionary articles. As constituents of articles data boxes do not form part of the obligatory micro-structure but rather enrich the obligatory microstructure, resulting in an extended obligatory microstructure. Data boxes in dictionary articles mostly contain text data but they can also display non-textual data. Data boxes are not default entries in any article but their use can be regarded as an extended treatment procedure. They can either contain items or item texts. Wiegand and Smit (2013: 153) distinguish as follows between items and item texts:

> An item is a functional text segment without the status of a sentence but with the status of a text segment which is given as a discernible item form assigned with at least one genuine function, the latter being precisely such that a user can obtain knowledge about the item's subject as lexicographical information.
>
> An *item text* is a functional text segment with item function and text constituent status exhibiting a complete and distinct natural-language syntactical structure and consisting of at least one sentence.

As constituents of dictionary articles data boxes may contain data presented in either condensed or non-condensed format. The articles in figure 1, a partial article stretch from the monolingual Afrikaans learner's dictionary *Basiswoordeboek van Afrikaans* (Gouws et al. 1994), (henceforth abbreviated as BW) contain two data boxes with items giving the pronunciation of the word represented by the lemma of the specific article. This dictionary only gives pronunciation guidance for a limited number of selected lemmata. It is not part of the obligatory microstructure and when pronunciation guidance is regarded as necessary, it is presented in a data box accommodated in the final slot of the article. The entries "Uitspraak 'gemie'" (= Pronunciation 'gemie') and "Uitspraak 'sjirurg'" (= Pronunciation 'sjirurg') are condensed forms and thus items.
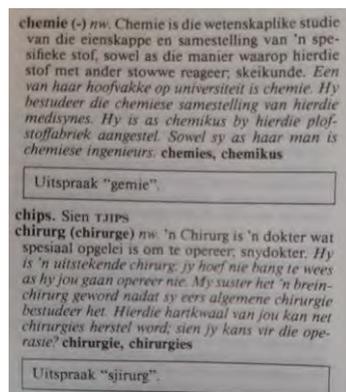
Figure 1:   Partial article stretch from BW

The data box in figure 2 from the *Junior Tweetalige Skoolwoordeboek/Junior Bilingual School Dictionary,* (Stoman et al. 2018), (henceforth abbreviated as JBS) contains a sentence which is an entry given in non-condensed form. This entry therefore is an item text.
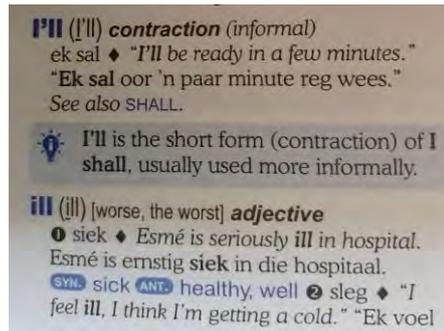
Figure 2:   Partial article stretch from JBS

Whether entries are items or item texts may have implications for the addressing relations in an article. This will be discussed in a further section of this article.

5.4.2   Positioning data boxes within dictionary articles

Data boxes are functional dictionary entries because they contribute to achieving the genuine purpose of the dictionary in which they occur. That implies that from the data on offer in a data box users must be able to retrieve information that can assist them in finding a solution to the problem that initiated the specific dictionary consultation procedure. Because data boxes result from extended treatment procedures and are usually not entries belonging to the obligatory microstructure of a dictionary, users will not know beforehand whether an article contains a data box. That is why it is important that these boxes need to be clearly marked as framed or coloured text constituents. Dictionary articles have no fixed slot reserved for data boxes. When planning the data distribution structure of a dictionary lexicographer's need to decide on the slots in dictionary articles that could accommodate search zones populated by data boxes. Data boxes could be placed in different text positions in dictionaries, cf. Taljard et al. (2014: 698). They could be included within the comment on form or comment on semantics or in an alternative position, e.g. as precomment or postcomment.

5.4.2.1 Data boxes at the end of an article

One of the most typical positions allocated to data boxes, cf. Gouws and Prinsloo (2010), is a slot at the end of the article. In such a position the data box often

falls beyond the scope of either the comment on form or the comment on semantics. As postcomment the data box is in a position of salience — further accentuated by its frame or colouring. Figure 3 shows the article of the lemma sign *especially* in the *Oxford Advanced Learner's Dictionary of Current English* (Turnbull 2010), (henceforth abbreviated as OALD):
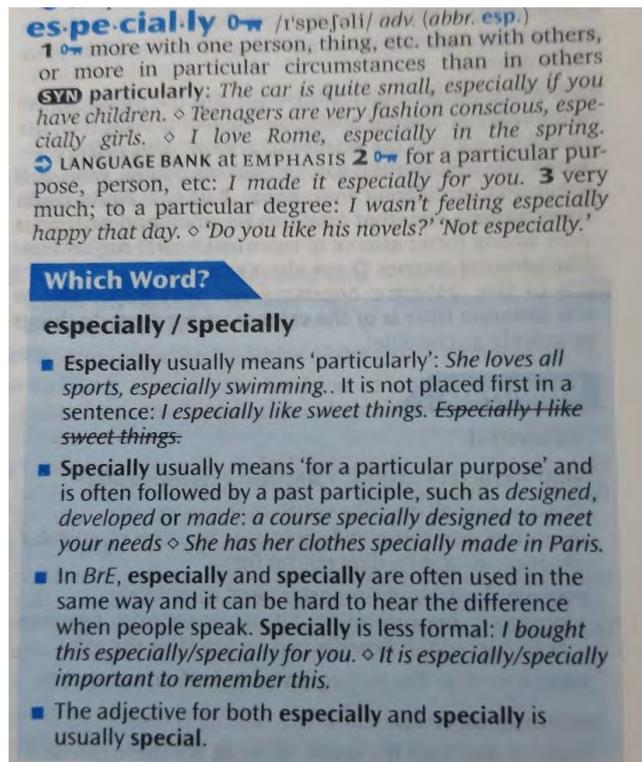
Figure 3:   Article in the OALD

The obligatory microstructure ends with the items in the third subcomment on semantics. The lexicographers have felt the need to present the users with additional guidance regarding the use of the words *especially* and *specially*. The data box containing this data has the title "Which Word?" and is presented after the comment on semantics as a postcomment to the article. Another example of a box presented as postcomment in the same dictionary can be found in the article of the lemma sign *restaurant*, figure 4. A type of data box frequently given in this dictionary contains collocations, but not necessarily collocations in which the word represented by the lemma functions as either base or collocator, but rather collocations applicable in the semantic domain of the word represented by the lemma sign. The user gets assistance regarding typical collocations used when dining in a restaurant:
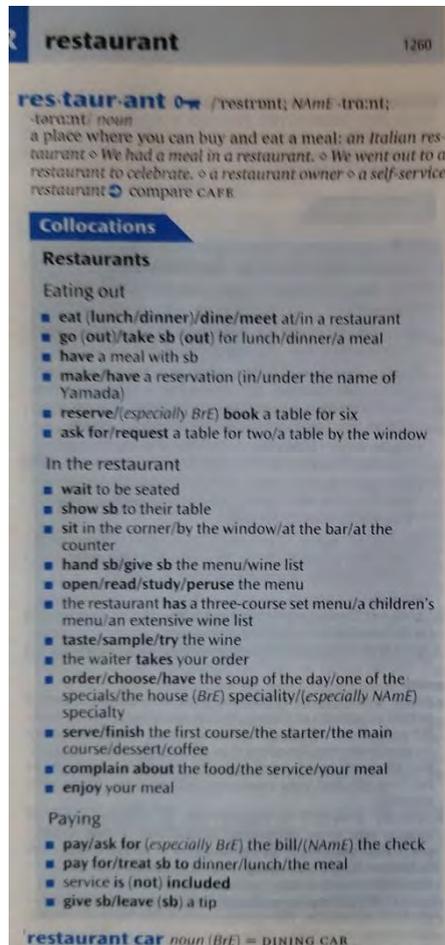
Figure 4:   Collocations in the OALD

The OALD presents the entries in data boxes against a blue background. These data boxes can easily be identified by the users of this dictionary.

There is no restriction on the number of data boxes per article, but lexicographers need to be careful not to have a data box overload that could diminish the emphasis on these boxes. An inflation of data boxes should therefore be avoided, but even more than one clearly marked data box per article should still be in order, as seen in figure 5, the article of the lemma sign *school* in the OALD. The two boxes "British/American" and "Grammar Point" are clearly identifiable and can enhance the nature and extent of the lexicographic treatment in this article.
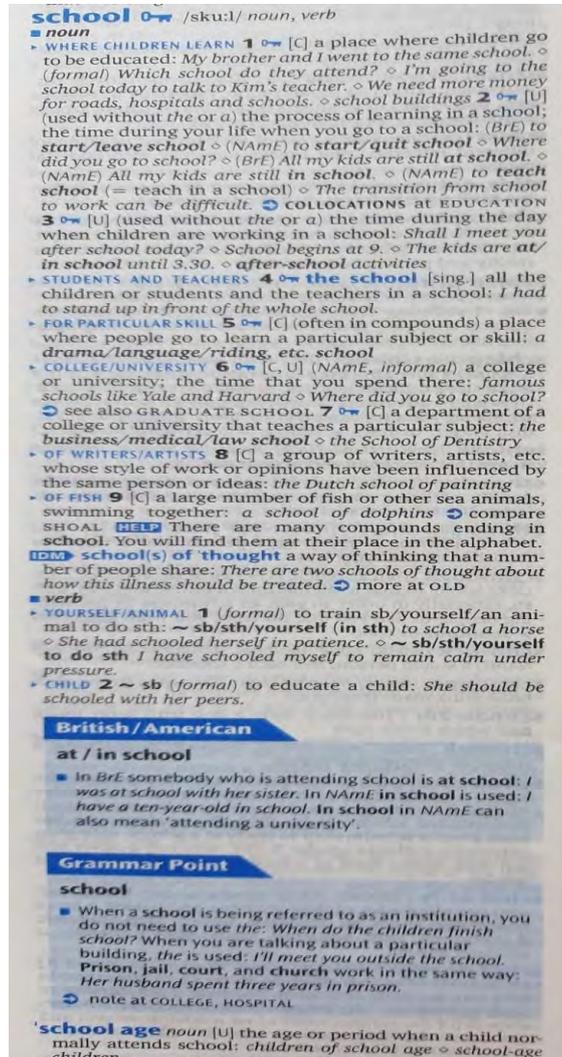
Figure 5:   Two data boxes in a single article of the OALD

Where a data box is presented as postcomment in the article of a lemma representing a polysemous lexical item it is not always clear whether the treatment presented in the box is directed at only one or more senses or at all the senses of the word. This problem will be elaborated on in section 5.4.2.4.

5.4.2.2 Data boxes at the beginning of the article

The lemma sign of an article is part of the main outer access structure of the

dictionary and a guiding element of the specific article. Irrespective of the type of information a user wants to retrieve from a dictionary article as search area the access to that item, especially in a printed dictionary, proceeds via the lemma sign onto the search paths of the inner access structure. Items positioned in search zones in close proximity of the lemma are in salient positions. The framed or coloured appearance of data boxes adds focus to their occurrence as text constituents of an article and if such a box is awarded a position close to the lemma sign it further elevates the salience of the data in the box.
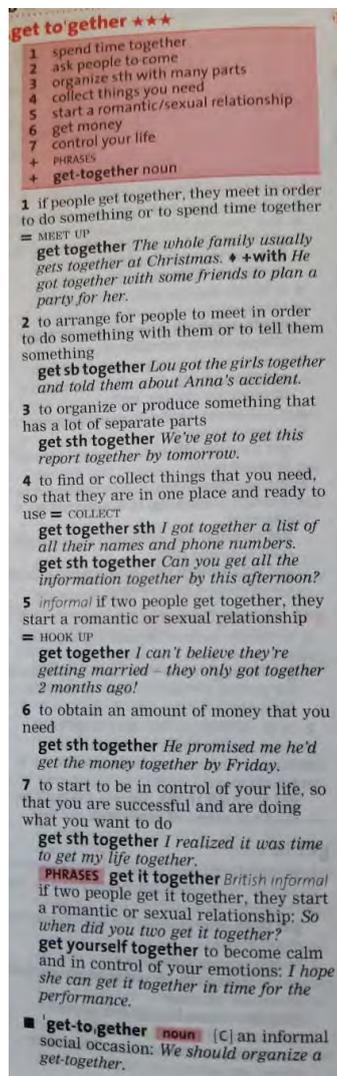
Figure 6:   The article of *get together* in the MPVP

The *Macmillan Phrasal Verbs Plus* (Rundell 2005), (henceforth abbreviated as MPVP) uses data boxes in a slot close to the beginning of an article, to provide an overview of the senses where the phrasal verb represented by the lemma sign has five or more than five senses. This box differs from other boxes because it does not contain the kind of data typically allocated to data boxes. It displays items conveying the menu of the comment on semantics and they function as navigational entries in this article. Alternative strategies for the provision of such navigational guidance could for instance be shortcuts, signposts and guidewords and will be discussed in detail in Part 2 of this series. The box as seen in figure 6 functions in such an article as a precomment.

Data boxes are also included as precomments of articles in the *Macmillan English Dictionary* (Rundell 2007), (henceforth abbreviated as MED). In figure 7, the article of the lemma sign *above*, the data box contains data regarding the use of the word *above* in its occurrence in different parts of speech. It makes the user aware of the differences between these uses.

Figure 7:   The article of *above* in the MED

To ensure that the target user of this learner's dictionary will focus on the differences between the uses of the word, this data can at best be presented in a precomment instead of being subsumed under the cotextual items in different comments or subcomments on semantics of one of the three partial articles of this twofold complex dictionary article, cf. Wiegand and Gouws (2011: 242).

Because the beginning of an article is a significant position of salience in a dictionary article lexicographers should carefully consider the type of data boxes to include in that search zone.

5.4.2.3 Data boxes as article windows

Lexicographers often have innovative ideas regarding the way of presenting data in their dictionaries. In the OALD several types of data boxes are used, with some types reserved for specific types of data. An article window is a type of data box reserved for a single type of data, i.e. word families, as seen in figure 8:
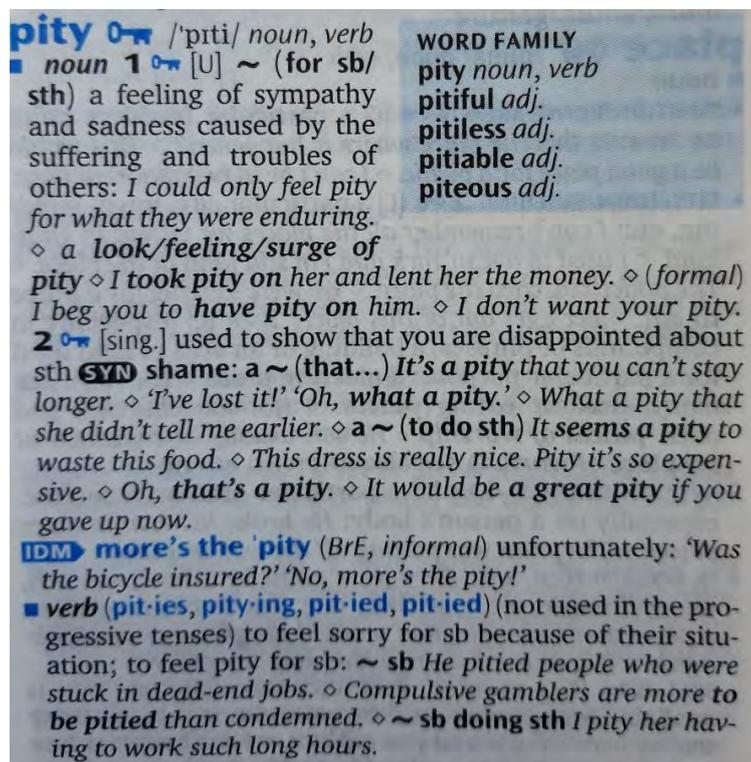
Figure 8:   An article window in the OALD

In the front matter text "Key to dictionary entries" it is said that: "Word families show words related to the headword." Although articles could in principle have windows in different article positions, e.g. top right, top centre, top left, middle right, middle centre, middle left, bottom right, bottom centre and bottom left, the OALD primarily uses the top right corner for its window data boxes. However, the layout of a dictionary article on the page could also have an influence on the position of the window. The editors of the OALD are consistent in always placing data boxes presented in article windows in the top right corner of a text block, albeit not necessarily of the article. Where an article

commences at the bottom of a column there could be a lack of space for a window in the top right corner of the article, as seen in figure 9.
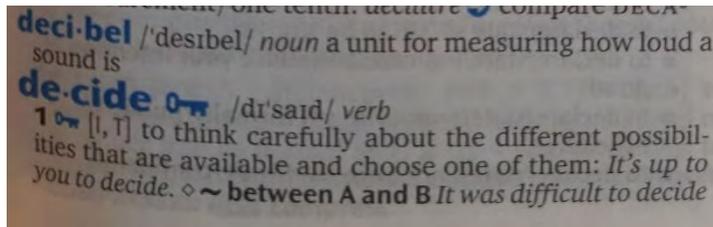
Figure 9:   Article in the OALD

However, the window is then positioned in the top right corner of the text block, containing the remainder of the article, that starts in a new column, as seen in figure 10.
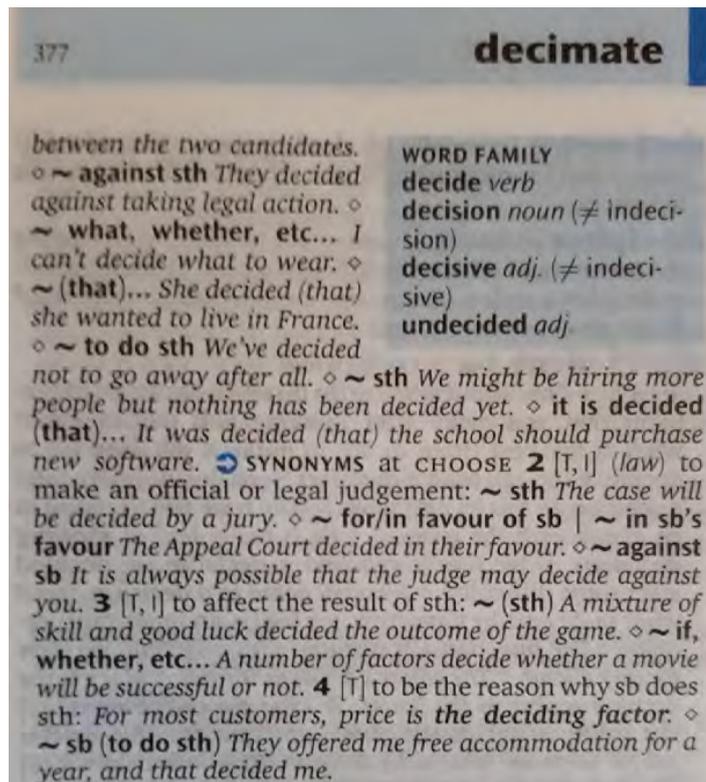
Figure 10:   Article window in the OALD

Technically such a window occupies the right centre of the article. From a lay-out perspective it could be argued that the OALD always presents its windowed data boxes in the top right corner of a text block, preferably but not always the top right corner of the article.

A brief identification of word families as found in the article windows can be valuable assistance to the target users of this dictionary. Users should be aware of this assistance but unfortunately the OALD does not provide an outer text in the front matter section with a list of articles that contain these windows. In spite of the focus that a data box puts on this data, there is no defined search path besides the main alphabetical outer access structure to guide users to these articles that contain article windows. This lack of predictable access impedes the added value of these article windows.

5.4.2.4 Data boxes within a comment of an article

Whether an article-internal data box contains an item or an item text, the contents of that data box is presented as part of a treatment procedure. This procedure is directed at one or more specific treatment units that constitute the address within this procedure of addressing. The positioning of article-internal data boxes directly reflects the relevant addressing relations in which data boxes can become involved.

Although lemmatic addressing, i.e. with the lemma as address, is the most frequently used addressing procedure in dictionaries, other types of addressing also prevail, especially non-lemmatic addressing. Various aspects of addressing are dealt with in Hausmann and Wiegand (1989), Louw and Gouws (1996), Gouws and Prinsloo (2005), Wiegand (2006; 2011) and Gouws (2014).
Lexicographers utilise data boxes to emphasise certain entries not accounted for by items in search zones of the obligatory microstructure of a dictionary. These items or item texts do not always have the lemma as address, but often occur within one of the comments of an article, especially the comment on semantics or a subcomment on semantics, where they participate in relations of non-lemmatic addressing. A less optimal article-internal positioning of a box of which the entry or entries are not addressed at all the senses of the word represented by the lemma sign of the article may diminish the added value the data box is supposed to have. In figure 11 from the *Handwoordeboek van die Afrikaanse Taal* (HAT) (Odendal and Gouws 2005) the data box is included as postcomment at the end of the article. The contents of the box only applies to the use of the word *instansie* in its second and third senses. From this presentation it is not clear that the box is not addressed at the first sense.
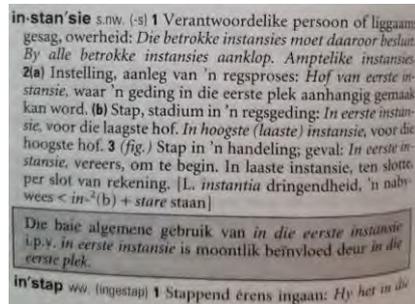
Figure 11:   Article of the lemma *instansie* in the HAT

Figure 12 is a partial article of the lemma *open²* from the *Longman Exams Dictionary* (Summers 2006), (henceforth abbreviated as LED). This article contains two data boxes, so-called study notes. Both these data boxes are presented within single subcomments on semantics for the first and the second sense of the polysemous word *open*. By positioning them there the lexicographer ensures that the user can know what the exact address of the data box is:
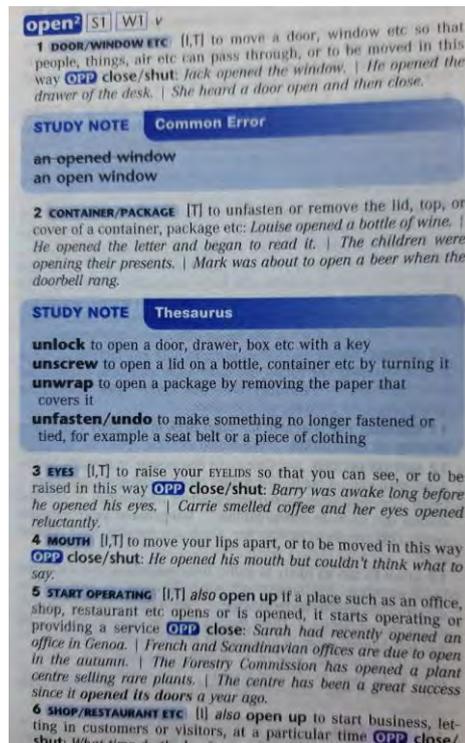
Figure 12:   Partial article of the lemma *open* in the LED

In the *Longman South African School Dictionary* (Bullon et al. 2007, (henceforth abbreviated as LSASD)) the article of the lemma sign *name¹* has three subcomments on semantics. The thesaurus data box is positioned at the end of the first subcomment on semantics. The thesaurus guidance in the data box is only applicable to the first sense of the word *name*.
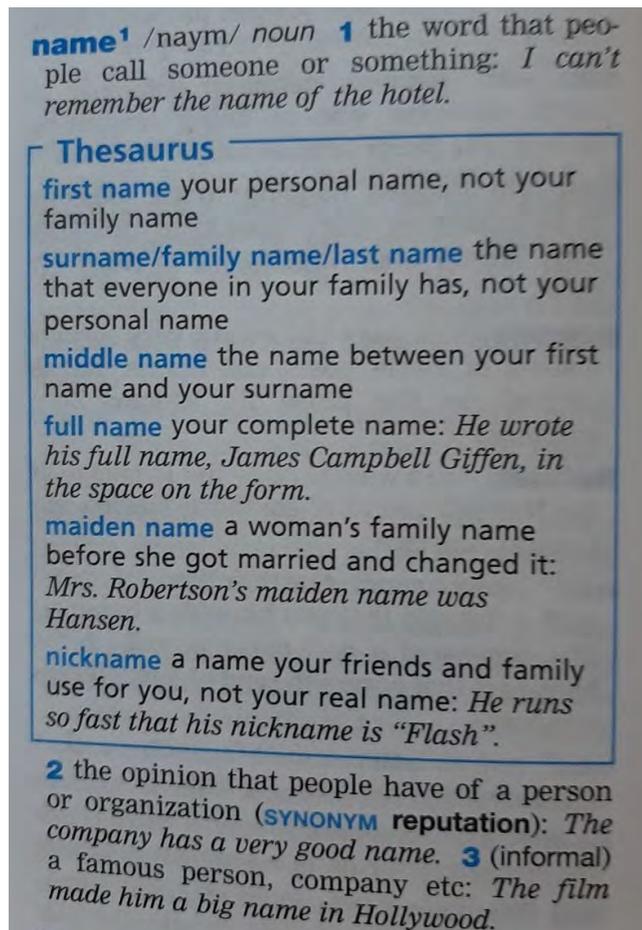
Figure 13: Article of the lemma *name¹* from the LSASD

Figure 14 from BW shows a data box that only has the first sense of the word *opmerking* as its address. Such a procedure of immediate addressing is to the benefit of the target user of the dictionary and demands less dictionary using skills compared to correctly interpreting the address of a data box that is not positioned in close vicinity to its address, even if the user was able to identify the appropriate sense:
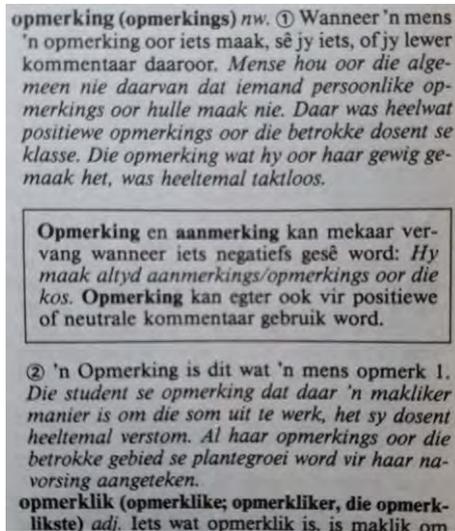
Figure 14:   Article of the lemma *opmerking* from BW

Where the treatment in a data box is directed at only one of the senses of a polysemous word or at only one of the items in the article and the data box is not positioned in such a way that an immediate addressing relation is possible, the lexicographer should clearly indicate what the address of the contents of the data box is. Such an approach is seen in figure 15, the article of the lemma *mad* in the *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary* (Pheiffer 2007), (henceforth abbreviated as OSD):
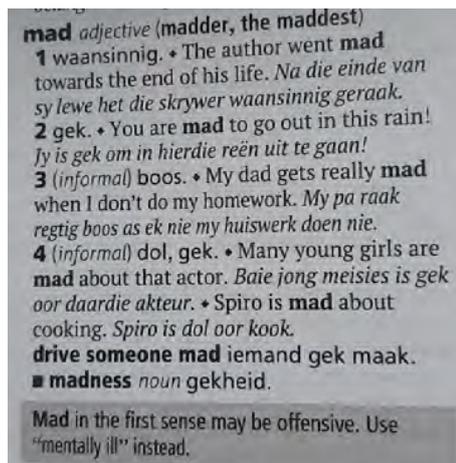
Figure 15:   Article of the lemma *mad* in the OSD

In this article with its four subcomments on semantics, the data box has been included as postcomment but it is clearly stated that it is addressed at the first sense.

Where a word can be used in more than one part of speech, the lexicographic treatment typically results in a complex article with a partial article for the occurrence of each of these parts of speech (Wiegand and Gouws 2011: 243). Data boxes could then be positioned in a search zone of one of the partial articles so that the other part of speech occurrence of the word falls beyond the scope of the data box. This can be seen in figure 16 from BW where the data box is positioned at the end of the first partial article of the article with the lemma *smaak* as guiding element.

Figure 16:   Article of the lemma *smaak* from BW

5.5   Phased-in inner texts as data boxes

5.5.1   Article-external text boxes

Data boxes can also be accommodated in the word list of dictionaries as phased-in inner texts. These data boxes are immediate constituents of the article stretches of the word list and function as article-external data carriers. Data boxes presented as phased-in inner texts typically have a connection with a lemma that is the guiding element of an article occurring in the specific article stretch into which the data box is phased in. These data boxes can be items or item texts and participate in either primary or secondary addressing relations. In addition to having the lemma that is in close proximity as address, these data boxes can also participate in relations of remote addressing, cf. Louw and Gouws (1996), where one or more lemmata in the same or often in other article stretches can be the address. As is the case with article-internal text boxes, different types of data can be presented in these data boxes, cf. Wiegand et al. (2017: 140-144 (WLWF)).

Phased-in inner texts could be constituents of the word list of a dictionary with a single alphabetical macrostructure or constituents of different word lists of dictionaries with poly-alphabetical macrostructures with vertical or horizontal parallel alphabetical access structures, cf. Wiegand (1989: 402) and Wiegand and Gouws (2013: 88). In the remainder of this section the focus will primarily be on phased-in inner texts in single macrostructures but reference will briefly be made to their occurrence in dictionaries with poly-alphabetical macrostructures with vertical parallel alphabetical access structures.

5.5.2   Phased-in inner texts in the primary macrostructure

In an expanded word list different types of phased-in inner texts can be included to split the article stretches.

5.5.2.1 Data boxes phased out of the article

In paragraph 5.4.2.3 it was mentioned that the layout of a dictionary page could influence the positioning of text boxes presented as article windows. In a comparable way the structure and layout of a dictionary article and the physical size of a specific item could lead to such an item not being included as microstructural constituent but rather in an article-external data box. This is where a phased-in inner text is the result of an item phased out of an article and phased into the article stretch as a data box.

Due to the position of an article in a column and on a page and due to their size pictorial illustrations do not always fit into the boundaries of an article as search area. The application of a well-defined data distribution structure

may then lead to a situation where such a pictorial illustration is phased out of the article and presented within a data box included as phased-in inner text.

The MWALD often contains pictorial illustrations as items in its articles, as can be seen in figure 17, the article of the lemma *glove*:
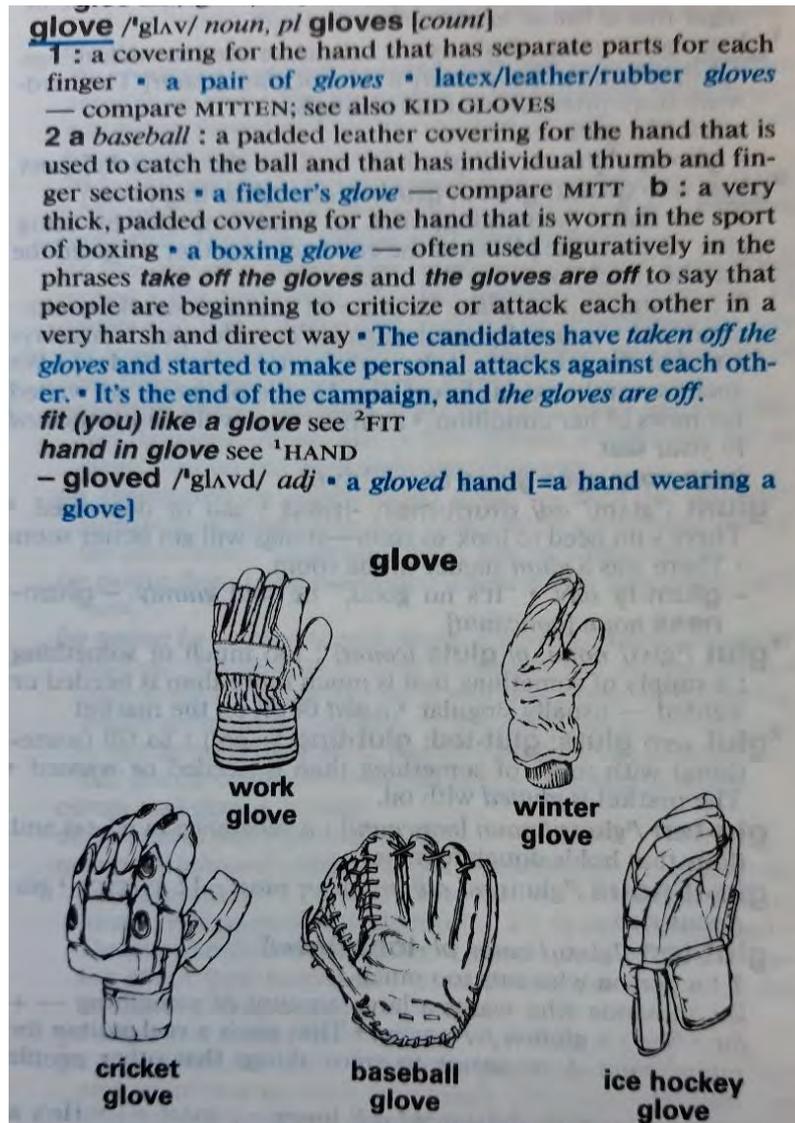
Figure 17:   Article of the lemma *glove* from the MWALD

No pictorial illustration is presented within the borders of the article of the lemma sign *deer* in the MWALD. However, an illustration does occur on the same page, in an article-external position, as can be seen in figure 18 where the illustration is entered across two columns and splits the partial article stretch between the articles of the lemmas *defamatory* and *defame*:
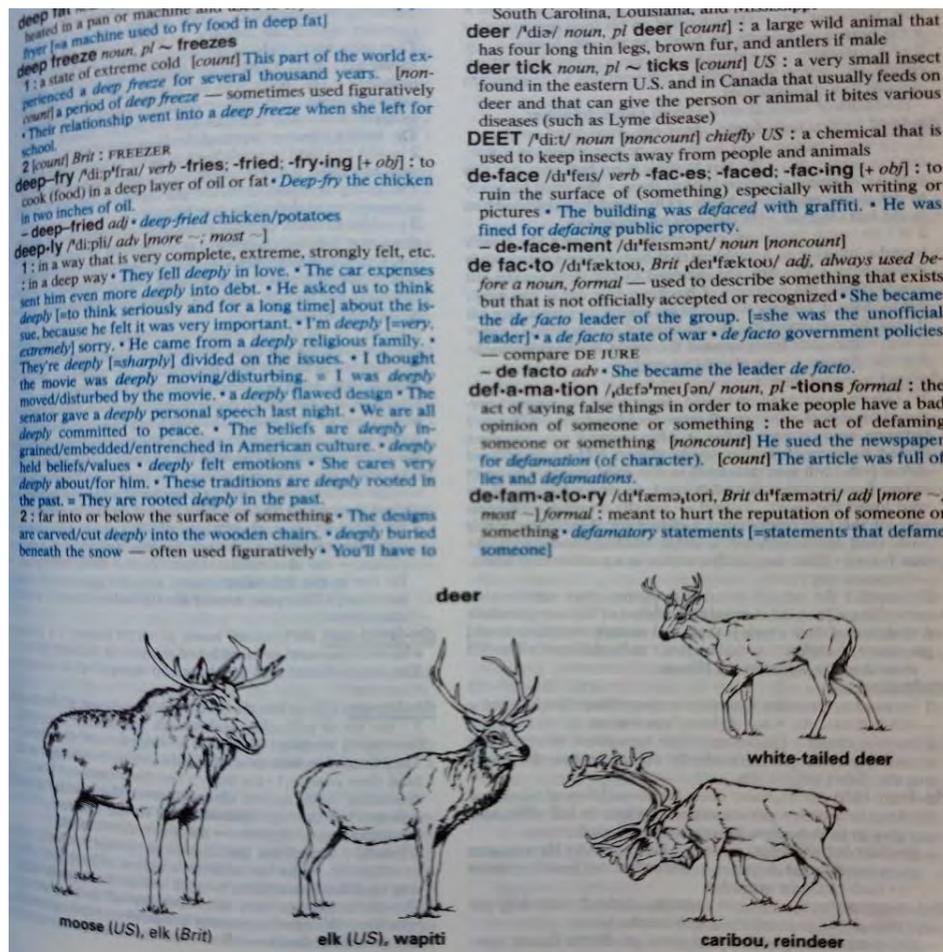
Figure 18:   Pictorial illustration in the MWALD

It could have been problematic for the lexicographers to fit this illustration into the article of the lemma *deer*. Consequently it has been phased out of the article and positioned as a phased-in inner text in the article stretch, where its title has the same form as the article of the lemma *deer*. Because the pictorial illustration

is in close proximity of the article of the lemma *deer*, this article has no reference to the pictorial illustration.[1]

In dictionaries a distinction can be made between single and synopsis articles, cf. Bergenholtz, Tarp and Wiegand (1999: 1780). The treatment in a single article is directed only at the lemma of that article, whereas the treatment in a synopsis article also contains data relevant for the treatment of the lemmata of other articles. Data boxes also have a single or a synopsis character. This applies to article-internal as well as article-external data boxes. Figure 19, the article for the lemma *lend* in the *Tweetalige aanleerderswoordeboek/Bilingual Learner's Dictionary* (Du Plessis 1993), (henceforth abbreviated as TAW), has an article-internal data box presented as postcomment. The contents of this data box apply to the lemma *lend* but it is also remotely addressed at the lemma *borrow*. Figure 20 shows the article of the lemma *borrow* with an item giving a cross-reference to *lend.*
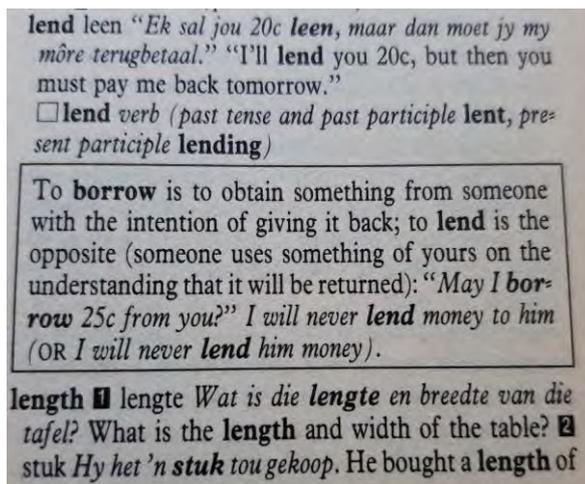
Figure 19:   Article of the lemma *lend* in the TAW
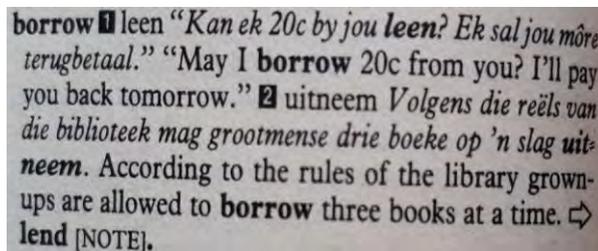
Figure 20:   Article of the lemma *borrow* in the TAW

The data box in the article of the lemma *lend* is a synopsis data box. In a similar way the data box titled *deer* is a synopsis data box. The pictures are also relevant for the treatment of lemmata like *moose, elk* and *caribou.* Consequently the articles of these lemmata have an item giving a cross-reference "see picture at DEER".

Positioning the contents of the data box *deer* not within the article but as an article-external phased-in inner text, can probably best be motivated on article and page layout grounds. A similar use of data boxes where the data have been phased out of an article and into a data box functioning as phased-in inner text, is found in the *Cambridge International Dictionary of English* (Procter 1995), (henceforth abbreviated as CIDE). Figure 21 shows the framed data box *Molluscs* presented as a phased-in inner text. The contents of this box could have been included in the nearby article of the lemma *mollusc.* The layout of the column might have been disturbed by this illustration and consequently it was phased-out into a data box in the partial article stretch but still in close vicinity of the lemma sign. This is also a synopsis data box that functions as the cross-reference address for cross-reference items in the articles of lemmata like *cuttle fish, octopus* and *snail*:
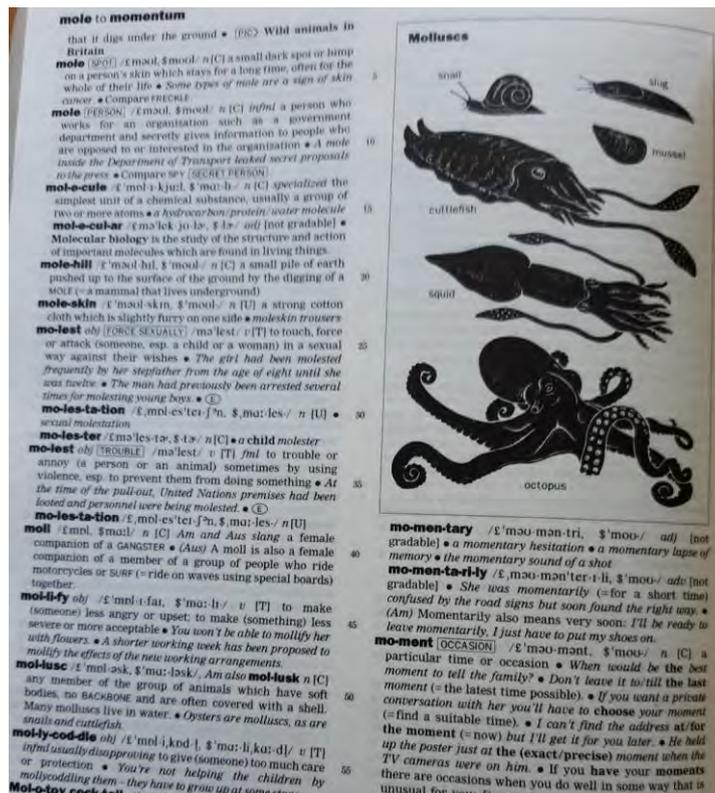
Figure 21:   Data box for *Molluscs* in the CIDE

Phasing-out procedures do not only target non-boxed data that could have been included in a search zone of an article and allocate them to an article-external data box. These procedures can also target data boxes that could have had an article-internal position and phase them out to function as article-external data boxes and phased-in inner texts. The OALD frequently includes a text box containing synonyms in a postcomment position of an article, as can be seen in figure 22, the article of the lemma *painful*:

Figure 22:   The article of the lemma *painful* from the OALD

This can be regarded as the default positioning of a data box that contains synonyms. The data box for synonyms is often phased out of the article and positioned as a data box and phased-in inner text in the textual position imme-

diately preceding the article. This is probably due to article and page layout considerations, a column break and the approach not to allow column or page breaks to divide a data box into two sections but always to present such a data box as an uninterrupted text block. This can be seen in figure 23, the article of the lemma *surprise:*

Figure 23:   Data box for the lemma *surprise* from the OALD

The biggest section of the article of the lemma *surprise* appears in the left column with a brief section of this article continuing in the right column. Had this section also been included in the left column there would probably not have been room for an uninterrupted data box as postcomment in that column. Consequently the data box was phased out to an article-external position, preceding the article. This data box is not positioned as precomment because it lies outside the article borders.

Because the planned data distribution structure of a dictionary needs to be executed in a meticulous and consistent way, it is not plausible to deviate from such a structure for column or page break or article or page layout reasons. Such a deviation impedes the access of users to the required data. The structural inconsistency resulting from the phasing out of items to article-external data boxes can be seen in the treatment of the lemma *save* in the OALD. In the OALD synonyms are frequently provided in data boxes, as seen in the treatment of *painful.* Irrespective of the degree of complexity of an article (whether it is a single or an n-fold complex article) or the number of senses treated in one or more subcomments on semantics, article-internal data boxes containing synonyms are presented as postcomments. Although such a positioning of the data box may cause addressing uncertainty in complex articles or articles with polysemous lexical items as lemmata, knowledgeable users will become familiar with this positioning of data boxes if this execution of the data distribution structure is done in a consistent way.

Column and page breaks result in deviations of this execution of the default data distribution structure. The treatment of the lemma *save* is presented in a threefold complex article with partial articles for the use of the word *save* as different parts of speech, namely as verb, noun, preposition and conjunction. The partial article treating the occurrence of *save* as a verb accounts for eight polysemous senses of this word, presented in different subcomments on semantics. The treatment is interrupted by a column and page break within the eighth subcomment on semantics. For the senses presented in the first two subcomments on semantics synonyms are provided in data boxes. The data box containing the synonyms for the first sense is included article-internally, according to the default data distribution of this type of text constituent, as a postcomment. However, the data box containing synonyms for the second sense has been phased out to a position immediately preceding the article of the lemma *save* as an article external phased-in inner text. This can be seen in figure 24 that shows the article as it is spread over the right column of the first and the left column of the second page:
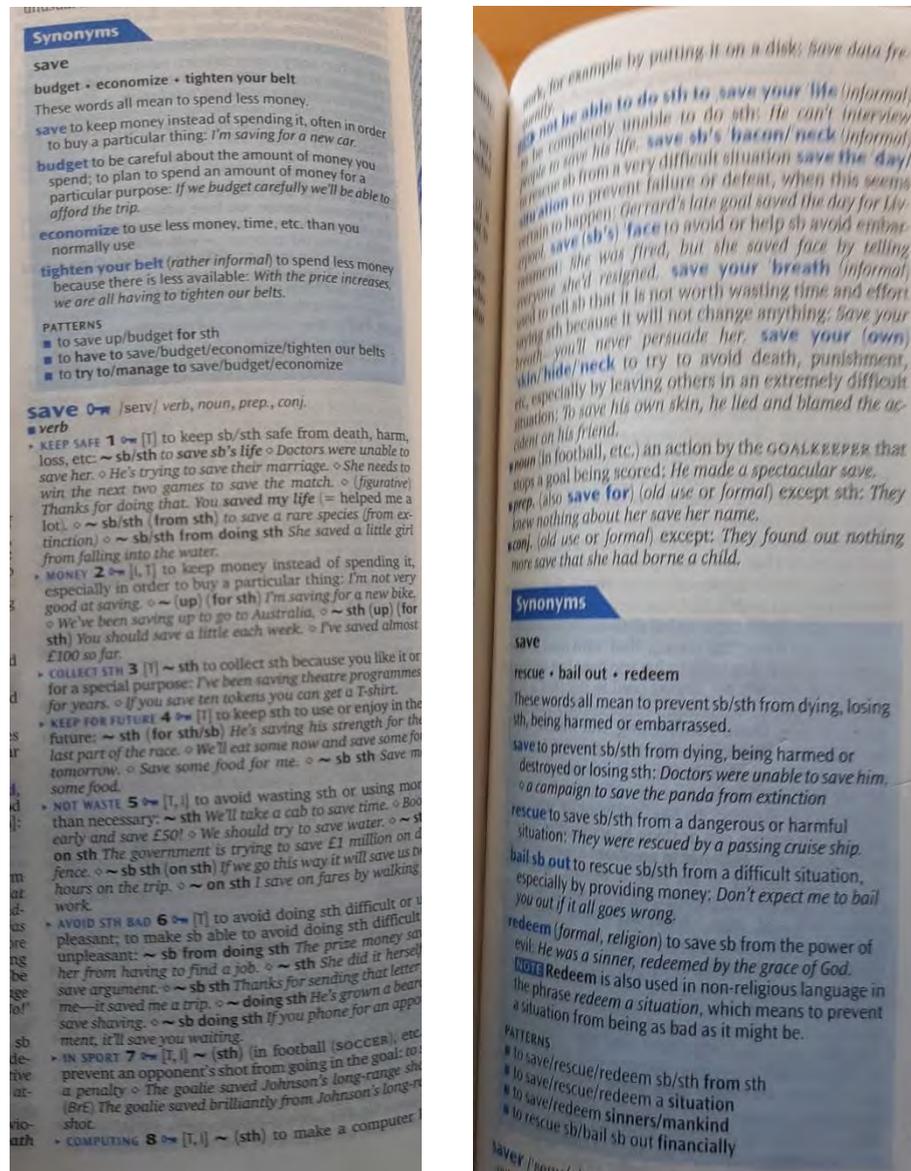
Figure 24:   Article of the lemma *save* from the OALD

Successfully accessing this data will be challenging for even a knowledgeable user of this dictionary. This is a typical situation where procedures in the lexicographic practice, i.e. allowing column and page breaks to impede the execution of the default data distribution structure, has resulted in theory-determined

dictionary structure problems. A more user-friendly way of dealing with these two data boxes would have been to either add them both as postcomments (reflecting the order of the senses at which they are addressed) or, even better, include each data box in a slot in the relevant subcomment on semantics. This would have enabled procedures of immediate addressing that would have assisted users in a quicker and better retrieval of information.

In some articles the inconsistency in the positioning of synonym data boxes in the OALD on column and page break grounds, results in such an article-internal data box not being included as the default postcomment but rather within the comment on semantics. The lemma *reason* has a single complex article with partial articles for its occurrence as noun and as verb. The article is interrupted by a column and page break, as seen in figure 25:
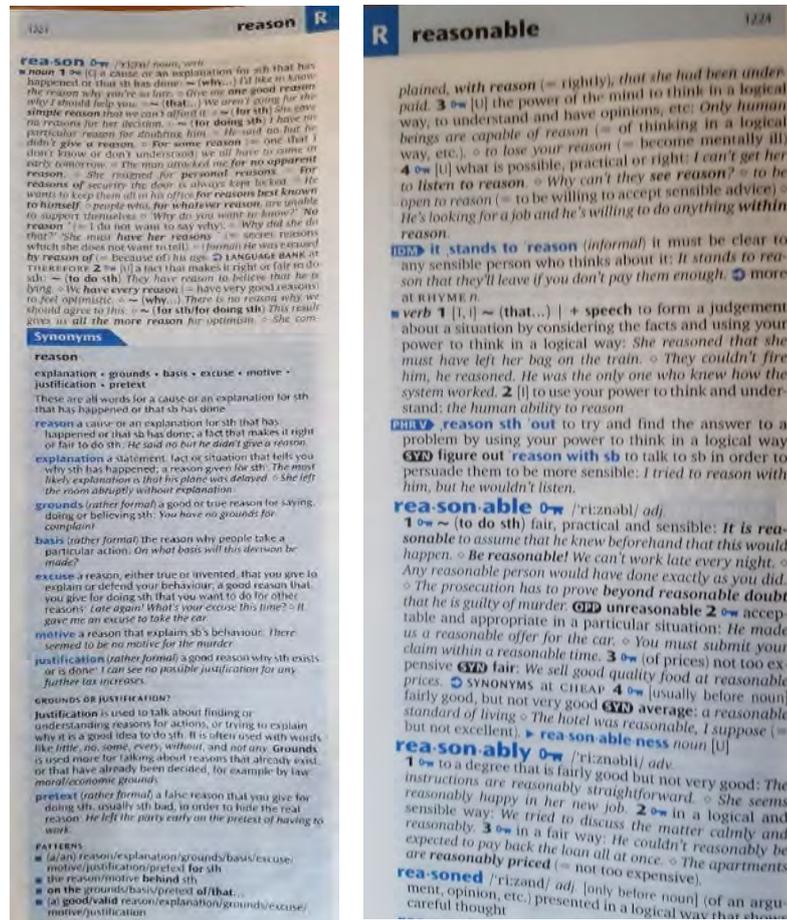
Figure 25:   Article of the lemma *reason* from the OALD

The comment on semantics of the partial article in which the occurrence as noun is treated, makes provision for four polysemous senses of this word. A data box is provided that contains synonyms for the first sense of this word. In order to present an uninterrupted text block the data box containing synonyms is positioned within the comment on semantics of the partial article in which the occurrence of *reason* as a noun is treated. It is positioned at the end of the first column on the first page on which the article is given. The data box occurs within the subcomment on semantics in which the second sense is treated, even though the synonyms are addressed at the first sense of the word. This positioning defies logical and consistent addressing relations.

5.5.2.2   Article-external data boxes resembling articles or article-internal items

Data boxes included in the word list as phased-in inner texts are often inserted between two other articles in a partial article stretch where it adheres to the alphabetical ordering. These data boxes typically have a guiding element that looks like a lemma and the box contains a rudimentary treatment of that word. In the TAW the word *fridge* had not been selected as a lemma candidate. However, the lexicographer must have felt the need to make the users of this learner's dictionary aware of that informal English word. Consequently, a data box was included as a phased-in inner text and inserted in its alphabetical position in the article stretch between the articles of the lemmata *Friday* and *friend*, as seen in figure 26, with the frame clearly identifying it as a data box and not a default article:
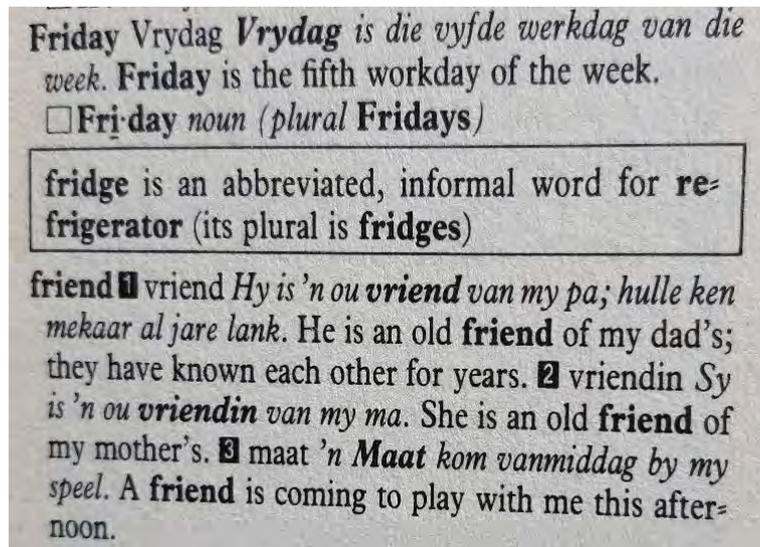
Figure 26:   A data box for *fridge* in the TAW

Figure 27, a partial article stretch from *Groot Woordeboek/Major Dictionary* (Eksteen et al. 1997); (henceforth abbreviated as GW), contains the data box with *percentage point* as its guiding element. This box follows the article niche attached to the article of the lemma *percentage.* This niche also contains the sublemma *percentage point* in a condensed format. In the niche the treatment of the lemma is restricted to an item giving a translation equivalent. The lexicographer wants to increase the assistance to users regarding this lemma and consequently employs the phased-in data box to supply the expression *percentage point* with a paraphrase of meaning.
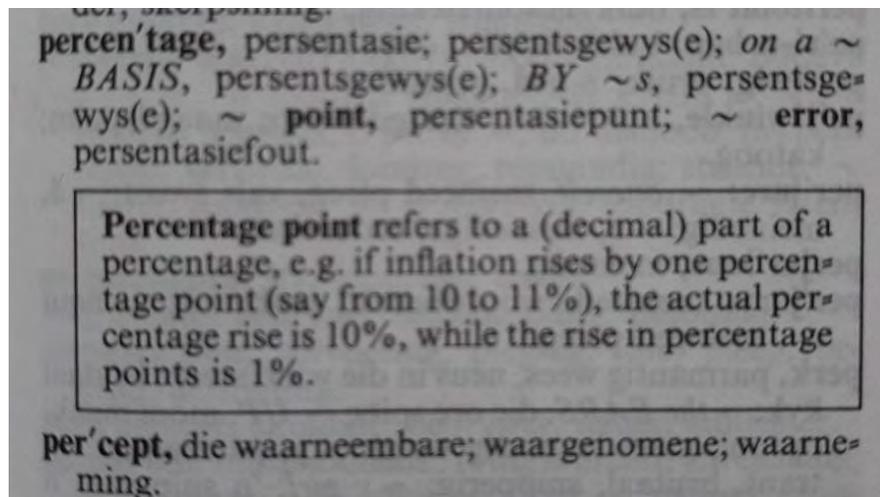
Figure 27:   Data box for percentage point from the GW

Data boxes do not always look like articles but can occur in the alphabetical position of a word that has not been included as lemma, in order to make the user aware of some relevant feature of that word. In Afrikaans the word *huidiglik* (=presently) is often used. However, from a linguistic perspective the use of this word is not approved. The word does not qualify for inclusion as a lemma, but the lexicographers of the monolingual Afrikaans school dictionary the *Oxford Afrikaanse Skoolwoordeboek* (Louw 2012), (henceforth abbreviated as OSW) would like to make their users aware of the fact that the word should not be used. The word *tans* should rather be used. In the alphabetical slot where *huidiglik* would have been entered had it been a lemma in this dictionary, the lexicographers include an article-external data box, cf. figure 28, to sensitise users that they should not use the word *huidiglik*.

huidige [hui·di·ge] *byvoeglike naamwoord* wat in
die teenwoordige tyd is, of wat nou gebeur:
*Wie is die **huidige** president van die VSA?*
◊ *Sy **huidige** salaris is nie baie hoog nie.*

Dit is verkeerd om van **huidiglik** te praat.
Gebruik eerder 'tans'.

huil *werkwoord* (**het gehuil**) **1** trane in jou oë
kry weens pyn of hartseer en soms ook kerm of
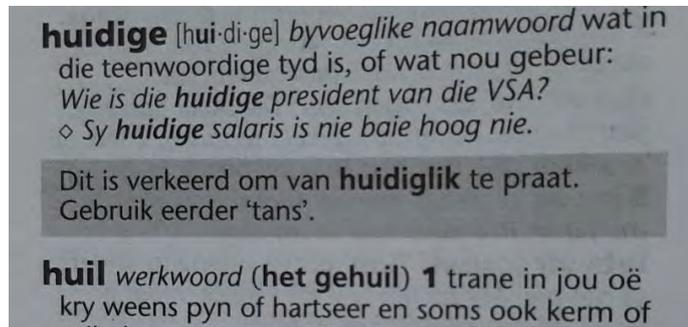
Figure 28:   Data box for *huidiglik* in the OSW

The data box in figure 28 is not a postcomment in the article of the lemma *huidige*
but rather a phased-in inner text.

Article-external data boxes often include data that do not represent a data
type that belongs to the default microstructural items but it does add to the treat-
ment of a given lemma or lemma cluster. Figure 29 presents a partial article
stretch from HAT:

hor·lo′sie *s.nw.* (-s; -tjie) ook **oorlosie 1** Toestel wat die tyd
aandui: *Wen die horlosie op; dit is afgeloop. Kyk op die horlosie
hoe laat dit is. Stel die horlosie reg. Sak-, arm-, pols-, staan-,
tafel-, muur-, hanghorlosie. 'n Horlosie met 'n wekker.* **2** *(in
die mv.)* Benaming wat in sekere streke vir die plant, elders
bekend as *turksnaels* en *muskusgras*, gebruik word. UITDR.:
*Met die een **oog** op die horlosie werk,* werk terwyl jy kort-kort
kyk hoe laat dit is om (a) op te hou of (b) betyds klaar te
wees. [G. *horologion* uurmelder < *hora* uur + *logein* sê] **hor-
losie: ~fabriek, ~veer, ~wyser.**

In samestellings met *horlosie-* as eerste stam kan *oorlosie-*
ook gebruik word. Die vorm met *horlosie-* is egter die ge-
bruiklikste en daarom die voorkeurvorm.

hor·lo′sie·band(jie) *s.nw.* Band(jie) waarmee 'n polshorlosie
om die arm vasgemaak word.

Figure 29:   Partial article stretch from the HAT

In the comment on form of the article of the lemma *horlosie* (watch) it is indi-
cated that *oorlosie* is a variant form of *horlosie*. Attached to the article of this
lemma is a lemma cluster, presented as a first level nest that includes a con-
densed presentation of compounds with *horlosie-* as first stem. Between this
nest and the following main lemma a data box has been inserted with guidance

regarding the formation of compounds with *horlosie-* as first stem. Although this data is relevant to the main lemma *horlosie-* and to the sublemmata presented in the nest, the data box is positioned outside the article of the main lemma as well as the nest of sublemmata and therefore it is a phased-in inner text.

5.5.2.3 Other article boxes included as phased-in inner texts

A variety of other types of data boxes also function as phased-in inner texts in dictionaries. A typical feature of many of these data boxes is their synopsis character whilst the data presentation in others complements and expands the default data coverage of a specific lemma. These data boxes convey data regarded by the lexicographer as important enough to include them in salient search venues like data boxes. The added value of these data boxes is unfortunately too often impeded by insufficient access guidance.

In close proximity of the article of the lemma *plural* in CIDE one finds the data box titled "Plural of nouns", presented in figure 30:

Figure 30:   Data box in the CIDE

The treatment in this data box is not addressed at any single lemma but it adds value to the treatment of all nouns that take plural. The problem is that it is very difficult for users in need of this guidance to know where to find the help. CIDE has a back matter text that provides an alphabetical index with relevant page numbers of all pictures, language portraits and lists of false friends included in the word list. The list includes an entry "Plurals" but no entry "Plural of nouns", the title of the data box in figure 30. It will remain a challenge for the target users of this dictionary to optimally benefit from this data box. They will probably only have access to this data box if they accidentally consult that page, seeing that the lemma *plural* is on the previous page and its article has no cross-reference to the data box.

Having consulted the lemmata *meet* and *meeting* in the CIDE a user has to turn the page to find the data box "Meeting someone", given in figure 31:
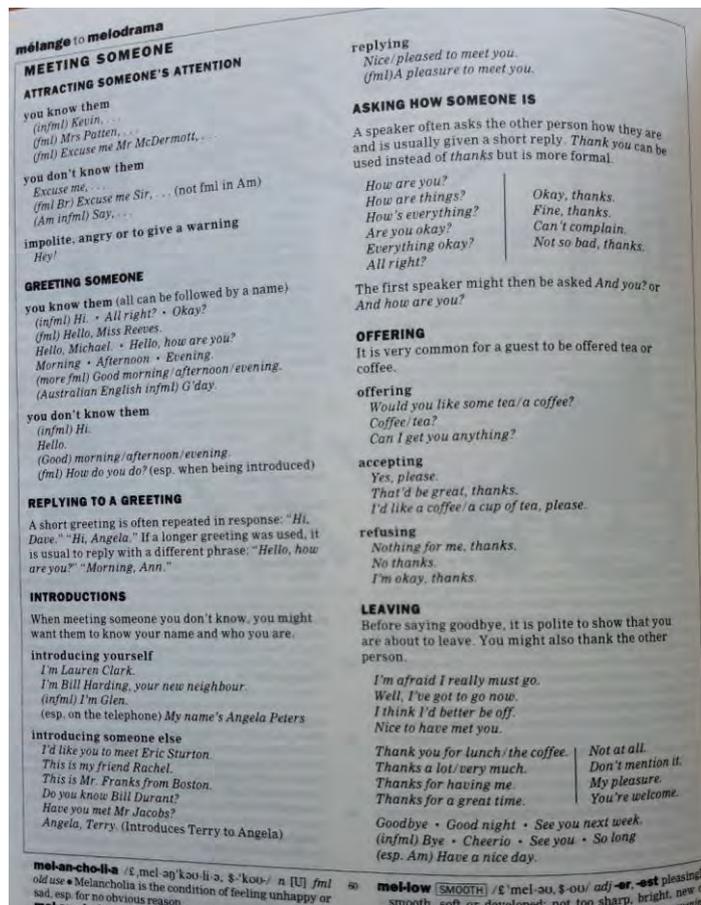
Figure 31:   Data box for *Meeting someone* from the CIDE

This data box contains pragmatic and cultural guidance that could assist the envisaged target users of this dictionary in their daily communication. Skilled users will be able to access this data box via the back matter list. However, for the user not aware of that list it would have been better if this data box could have been positioned on the same page where the lemmata *meet* and *meeting* appear.

On the same page in CIDE where the lemma *dash* is treated a data box has been presented that offers valuable complementary assistance regarding the use of a dash, cf. figure 32:

Figure 32:   Data box for *dash* from the CIDE

The contents of a data box like this one can hardly be included in the article of the corresponding lemma. Data boxes like these elevate the quality of the lexicographic treatment.

Even if an article in a learner's dictionary displays an extended obligatory microstructure lexicographers have to restrict the extent of the data presented in the different search zones and should refrain from procedures of data overload. If lexicographers want to respond to the needs of the users of a learner's dictionary, like CIDE, for some linguistic guidance typically contained in text books, they can use data boxes included as phased-in inner texts to convey this kind of data. Figure 33, a partial article stretch from CIDE, contains articles of the lemmata *homograph* and *homophone.* In each one of these articles there is an item giving a cross-reference to the language portrait (LP) *Homophones and homographs.*
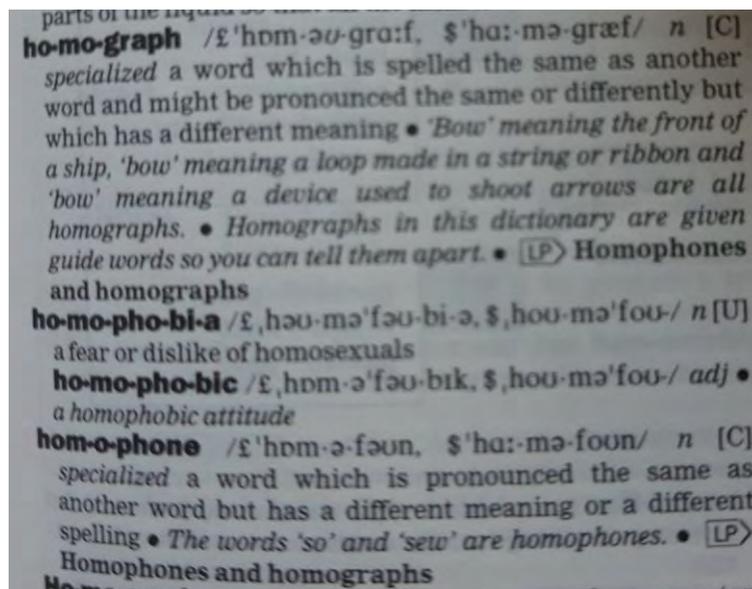
Figure 33:   Partial article stretch from the CIDE

This language portrait is presented as the data box seen in figure 34. This data box can also be accessed via the index in the back matter text. It has a synopsis character that could assist users in clearly distinguishing between homophones and homographs.

Figure 34:   Data box from the CIDE

5.5.3   Phased-in inner texts in the secondary constituent of a parallel macro-structure

Dictionaries can have more than one word list and therefore more than one macro-structure. More than one macrostructure could prevail within the same alphabetical constituent of a dictionary. One such type identified by Wiegand (1989: 402) and Wiegand and Gouws (2013: 88) is the poly-alphabetical macrostructures

with vertical parallel alphabetical access structures. The *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary* (Grobbelaar 1987), (henceforth abbreviated as RWD), displays these macrostructures. The primary macrostructure of this dictionary is presented in two central columns on each page. An additional column on the left and an additional column on the right of the two central columns constitute a secondary macrostructure. The partial article stretch of the secondary macrostructure on each page falls within the alphabetical domain of the partial article stretch of the primary macrostructure on the same page. The RWD is an expanded version of a previously published bilingual dictionary. The primary macrostructure is an unchanged version of that of the previous dictionary. The secondary macrostructure is constituted by new text constituents. In the secondary word list there are articles that have a selection of lemmata from the primary macrostructure as guiding elements. A different and innovative treatment (not to be discussed here) has been executed in these articles. The secondary macrostructure displays an expanded word list with partial article stretches split by phased-in inner texts, as seen in figure 35:

Figure 35:   Poly-alphabetical macrostructure from the RWD

The text boxes presented as phased-in inner texts are titled *Words in action*. They are addressed at some of the source and target language items in the articles of the primary macrostructure and they contain different types of data, e.g. pragmatic, usage, linguistic and lexical guidance.

The data boxes help to introduce new comments to enhance the quality of the original dictionary.

Conclusion

Data boxes occur frequently and in diverse ways in dictionaries. This is especially the case in printed dictionaries, although some electronic dictionaries, especially those that are based on printed dictionaries, also employ this type of lexicographic entry. Although various aspects of the use of data boxes have been discussed in metalexicographic literature, much still needs to be done in this regard. This paper focused on what data boxes are, identified them as a type of lexicographic entry and indicated where they are positioned in dictionaries and that they should be used in such a way that they add value to dictionaries. This can serve as an aid for future lexicographers who wish to employ data boxes in their dictionaries.

Endnote

1.   The value of a picture presenting different types of deer (but not all types) is not discussed here. The decision to give a pictorial illustration rests with the lexicographer. Therefore it cannot be expected that a specific word will be illustrated or illustrated in the same way in all dictionaries. For example, the online version of the MED does not contain such an illustration.

Acknowledgement

References

Dictionaries

BW = Gouws, R., I. Feinauer and F. Ponelis. 1994. *Basiswoordeboek van Afrikaans.* Pretoria: J.L. van Schaik.

CIDE = Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English.* Cambridge: Cambridge University Press.

GW = Eksteen, L.C. et al. (Eds.). 1997[14]. *Groot Woordeboek/Major Dictionary*. Cape Town: Pharos.

HAT = Odendal, F.F. and R.H. Gouws. 2005[5]. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Cape Town: Pearson Education South Africa.

JBS = Stoman, A. et al. (Eds.). 2018. *Junior Tweetalige Skoolwoordeboek/Junior Bilingual School Dictionary.* Cape Town: Pharos.

LED = Summers, D. (Ed.). 2006. *Longman Exams Dictionary.* Harlow: Pearson Education.

LSASD = Bullon, S. et al. (Eds.). 2007. *Longman South African School Dictionary.* Harlow: Pearson Education.

MED = Rundell, M. (Ed.). 2007[2]. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

MPVP = Rundell, M. (Ed.). 2005. *Macmillan Phrasal Verbs Plus.* Oxford: Macmillan Education.

MWALD = Perrault, S.J. (Ed.). 2008. *Merriam-Webster's Advanced Learner's English Dictionary*. Springfield, Massachusetts: Merriam-Webster.

OALD = Turnbull, J. (Ed.). 2010[8]. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.

OSD = Pheiffer, F. et al. (Eds.). 2007. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. Cape Town: Oxford University Press Southern Africa.

OSW = Louw, P. (Senior Editor). 2012. *Oxford Afrikaanse Skoolwoordeboek.* Cape Town: Oxford University Press Southern Africa.

RWD = Grobbelaar, P. 1987. *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary*. Cape Town. The Reader's Digest Association, South Africa (Pty) Ltd.

TAW = Du Plessis, M. 1993. *Tweetalige aanleerderswoordeboek/Bilingual Learner's Dictionary.* Cape Town: Tafelberg.

WLWF = Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.). 2017. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research.* Vol. 2. Berlin/New York: Walter de Gruyter.

Other references

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2*: 1762-1832. Berlin: De Gruyter.

Gouws, R.H. 2014. Expanding the Notion of Addressing Relations. *Lexicography: Journal of ASIALEX* 1(2): 159-184.

Gouws, R.H. 2018. Expanding the Data Distribution Structure. *Lexicographica* 34: 225-237.

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS, AFRICAN SUN MeDIA.

Gouws, R.H. and D.J. Prinsloo. 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010:* 501-511. Leeuwarden: Fryske Akademy.

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

Gouws, R.H. et al. (Eds.). 2013. Dictionaries. *An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* Berlin/New York: De Gruyter.

Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. 1989–1991: 328-360.

Hausmann, F.J. et al. (Eds.). 1989–1991. *Worterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlin: Walter de Gruyter.

Louw, P.A. and R.H. Gouws. 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.

Taljard, E., D.J. Prinsloo and R.H. Gouws. 2014. Text Boxes as Lexicographic Device in LSP Dictionaries. Abel, A., C. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the 16th EURALEX International Congress*: *The User in Focus, Bolzano/Bozen, Italy, 15–19 July 2014*: 697-705. Bolzano/Bozen: EURAC.

Tarp, S. 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.

Wiegand, H.E. 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. 1989–1991: 371-409.

Wiegand, H.E. 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II*: 1-82. Tübingen: Max Niemeyer.

Wiegand, H.E. 2006. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 22: 187-261.

Wiegand, H.E. 2011. Adressierung in der ein- und zweisprachigen Lexikographie. Eine zusammenfassende Darstellung. Kürschner, W. and M. Ringmacher (Eds.). 2011. *Aus Ost und West:* 109-234. Frankfurt: Peter Lang.

Wiegand, H.E., S. Beer and R.H. Gouws. 2013. Textual Structures in Printed Dictionaries. An Overview. Gouws, R.H. et al. (Eds.). 2013: 31-73.

Wiegand, H.E. and R.H. Gouws. 2011. Theoriebedingte Wörterbuchformprobleme und wörterbuchformbedingte Benutzerprobleme I: Ein Beitrag zur Wörterbuchkritik und zur Erweiterung der Theorie der Wörterbuchform. *Lexikos* 21: 232-297.

Wiegand, H.E. and R.H. Gouws. 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 273-314.

Wiegand, H.E and M. Smit. 2013. Microstructures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 149-214.

Lexicographic Data Boxes.
Part 2:
Types and Contents of Data Boxes with Particular Focus on Dictionaries for English and African Languages[*]

D.J. Prinsloo, *Department of African languages, University of Pretoria, South Africa (danie.prinsloo@up.ac.za)*
and
Rufus H. Gouws, *Department of Afrikaans and Dutch, Stellenbosch University, South Africa (rhg@sun.ac.za)*

Abstract: This article, the second in a series of three on lexicographic data boxes, focuses primarily on the types and contents of data boxes with particular reference to dictionaries for English and African languages. It will be proposed that data boxes in paper and electronic dictionaries can be divided into three categories and that a hierarchy between these types of boxes can be distinguished, i.e. (a) a bottom tier — data boxes used as mere alternatives to other lexicographic ways of presentation such as the bringing together of related items and/or to make entries visually more attractive, (b) a middle tier — addressing more salient features e.g. range of application, contrast, register, restrictions, etc. and (c) a top tier — vital salient information, e.g. warnings, taboos and even illegal words. A distinction is made between data boxes which are universal in nature, i.e. applicable to any language, data boxes pertaining to a language family and data boxes applicable to a specific language.

Keywords: DICTIONARIES, LEXICOGRAPHIC DATA BOXES, TEXT BOXES, SHADED BOXES, AFRICAN LANGUAGES, SEPEDI, ISIZULU

Opsomming: Leksikografiese datakassies. Deel 2: Tipes datakassies en hulle inhoud, met spesifieke verwysing na woordeboeke vir Engels en die Afrikatale. Hierdie artikel, die tweede in 'n reeks van drie oor leksikografiese datakassies, fokus hoofsaaklik op die tipes datakassies en hulle inhoud, met spesifieke verwysing na woordeboeke vir Engels en die Afrikatale. Daar sal voorgestel word dat datakassies in papier- en elektroniese woordeboeke in drie kategorieë verdeel kan word en dat 'n hiërargie tussen hierdie tipes kassies onderskei kan word, d.w.s. (a) 'n onderste vlak — datakassies wat slegs as alternatiewe vir ander leksiko-

---

[*] This is the second in a series of three articles dealing with various aspects of lexicographic data boxes.

grafiese aanbiedingsmetodes gebruik word soos die bymekaarbring van verwante items en/of om inskrywings visueel aantrekliker te maak; (b) 'n middelvlak — om meer opvallende kenmerke aan te spreek, bv. die reikwydte, kontras, register, beperkings, ens. en (c) 'n hoogste vlak — essensiële inligting, bv. waarskuwings, taboes en selfs onwettige woorde. Daar word onderskei tussen data-kassies wat universeel van aard is, dit wil sê van toepassing op enige taal, datakassies wat ter sake is vir 'n taalfamilie en datakassies wat van toepassing is op 'n spesifieke taal.

Sleutelwoorde: WOORDEBOEKE, LEKSIKOGRAFIESE DATAKASSIES, TEKSKASSIES, SKADU-DATAKASSIES, AFRIKATALE, SEPEDI, ISIZULU

1.   Introduction

Data boxes are commonly used in paper and electronic dictionaries to convey a variety of data not typically catered for by, what could be called standard presentation procedures that employ for example items giving the paraphrase of meaning (definitions), translation equivalents, examples of usage, pictorial illustrations, pronunciation guidance, and frequency indicators. Data boxes are used in cases where data entries are required to improve the lexicographic presentation and treatment — they add value to the default treatment. They typically include a variety of data types such as guidance in terms of grammar, pronunciation, sense distinction, contrasting related words, restrictions on the range of application, register, pronunciation, etc.

Gouws and Prinsloo (2005: 133) state that:

> ... text boxes are put to good use to convey relevant data which falls outside the scope of the default categories presented in the normal search fields of the article.

Oxford Bilingual School Dictionary: Zulu and English (OZSD) and *Oxford Bilingual School Dictionary: Northern Sotho and English* (ONSD) refer to their shaded boxes as usage notes and describe their nature as follows.

> Usage notes guide learners on potential areas of difficulty, helping them avoid common mistakes. Usage notes are also used to give additional information on how and when to use a headword (OZSD and ONSD: vi).

In the section "using your dictionary", *Macmillan English Dictionary for Advanced Learners* (MED) distinguishes between three types of shaded boxes, i.e. "information to learn more about how a word is used", "hints to avoid common errors" and notes that tell you about the origin of a word". *Oxford Advanced Learner's Dictionary of Current English* (OALDC) (Appendix 9: 1414) provides notes on usage of various types, e.g. clarification of grammar aspects, British and American usage or dealing with differences between words with similar meanings. *Reader's Digest Afrikaans–Engelse Woordeboek / English–Afrikaans Dictionary* (RWD) (page 5) informs the user about shaded boxes announced as "understand the other language as never before".

> … there are always problems that constantly trip one up. In order to help you overcome the trickier points of style and usage we have included hundreds of 'words in action' …

However, in spite of the frequent occurrence of data boxes in a variety of dictionary types, relatively little has been done to analyse data boxes with regard to the data types included in these boxes or the typological range of data boxes. This article embarks on an effort to identify different types of data found in data boxes of existing paper and electronic dictionaries and suggests that these boxes can be divided into three categories based upon type and content. It will be proposed that a hierarchical ordering between these categories can be distinguished, i.e. (a) a bottom tier — data boxes used as mere alternatives to other lexicographic ways of presentation, e.g. mere groupings or bringing together of related items. This is often done to make an entry visually more attractive; (b) a middle tier — giving more data, comparable to the type of additional data often found through cross-references, but addressing more salient features and (c) a top tier — vital salient data, e.g. warnings, taboos and even illegal words. Any attempt at the classification of data boxes is, however, arbitrary — no water tight classification is possible since a single data box often deals with a variety of issues as in figure 1. This data box primarily displays words and expressions semantically related to the word *mad*, but it also conveys other types of usage guidance. A number of bullets deal with register, i.e. formal versus informal use of the word, the third and fifth bullets deal with offensive use, the sixth bullet gives grammatical restrictions, and bullets 2, 3, 4, and 7 contrast language variations i.e. British English versus American English in this case.

IDIOMS

like crazy/mad (INFORMAL) very fast, hard, much, etc
▪ *We worked like crazy to get it done on time.*

✖ **Usage note:** mad

crazy nuts batty out of your mind (not) in your right mind

- These are all informal words that describe somebody who has a mind that does not work normally.
- **mad** (*informal, especially British English*) having a mind that does not work normally: *I thought I'd go mad if I stayed any longer.*
- **Mad** is an informal word used to suggest that somebody's behaviour is very strange, often because of extreme emotional pressure. It is offensive if used to describe somebody suffering from a real mental illness; use **mentally ill** instead. **Mad** is not usually used in this meaning in North American English; use **crazy** instead.
- **crazy** (*informal, especially North American English*) having a mind that does not work normally: *A crazy old woman rented the upstairs room.*
- Like **mad**, **crazy** is offensive if used to describe somebody suffering from a real mental illness.
- **nuts** [not before noun] (*informal*) mad: *That noise is driving me nuts!* ◇ *You guys are nuts!*
- **batty** (*informal, especially British English*) slightly mad, in a harmless way: *Her mum's completely batty.*
- **out of your mind** (*informal*) unable to think or behave normally, especially because of extreme shock or anxiety: *She was out of her mind with grief.*
- **(not) in your right mind** (*informal*) (not) mentally normal: *No one in their right mind would choose to work there.*
- to be **mad/crazy/nuts/out of your mind/not in your right mind to do something**
- to **go mad/crazy/nuts/batty**
- to **drive somebody** mad/crazy/nuts/batty/out of their mind
- **completely** mad/crazy/nuts/batty/out of your mind

Figure 1:   Data box at *mad* in *Oxford Advanced Learner's Dictionary of Current English* (OALD)
http://oald8.oxfordlearnersdictionaries.com/dictionary/crazy

Different scopes of application can also be distinguished, i.e. data box types which are (a) general in nature and not restricted to any specific language; (b) data box types pertaining to a language family and (c) data box types applicable to a specific language. Typical examples of the general utilization of data boxes are specifying semantic and syntactic restrictions, contrasting related words, warning against improper use, etc. Data boxes applicable to a language family deal with data that members of a specific language family have in common. Typical examples, given in a next section, of data boxes pertaining to a language family are those dealing with nominal classes, concords and pronouns. For a specific language it would, e.g. be data boxes giving syntactic restriction for specific words, e.g. question particles *afa* and *afaeya* in Sepedi.

This article does not take a critical approach to either the contents and presentation of data boxes or whether a specific entry that might perhaps be regarded as a data box in a current dictionary actually qualifies to be called a data box. Criteria for data boxes have yet to be formulated and it will not be done in this article. Data boxes are typically presented as frames or as a coloured background to one or more items in a dictionary. For the purpose of this article the occurrence of frames as a slot for the accommodation of certain items or of a coloured section functioning as highlighting background to certain items will be regarded as data boxes. A critical assessment with proposals for what should actually qualify as a data box is envisaged for the last article in this trilogy.

A topic not discussed in this article regards the metalanguage used in data boxes in bilingual dictionaries. Arguments could be offered that the metalanguage should be the source language of a monodirectional or of a specific component of a bidirectional bilingual dictionary, but equally compelling arguments could be offered that it should be the target language in both these dictionary types. The decision regarding the metalanguage should not be done in a haphazard way. Lexicographers need to determine the needs and reference skills of their target users and the lexicographic functions to be satisfied by a given dictionary. These matters should be considered when making a decision regarding the metalanguage to be used in the data boxes of any given bilingual dictionary, but space constraints do not allow a full investigation into this aspect in this article.

Updating both printed and online dictionaries inevitably leads to changes that can also influence their use of data boxes. The data boxes discussed in this paper come from specific editions and versions of the respective printed or online dictionaries. Some of these data boxes no longer appear in the most recent editions or versions. The authors of the article are aware of this situation but still use these examples due to their applicability to the discussion of specific contents or type of data box.

2.   Proposed hierarchy of data boxes as found in current dictionaries

2.1   A bottom tier of data boxes

In this category data boxes are utilized for mere groupings, bringing together of related items, and to make entries visually more attractive. The first type of data box in the bottom tier that could be distinguished is a box containing a list that *brings together the different senses* in a menu that provides a quick overview, as in figure 2 in MED.

Figure 2:   Boxed menu of senses for *sharp* in MED

The boxed senses in figure 2 could as well be presented in an alternative way, consider the same lemma in the paper version versus the *Macmillan Dictionary* (OMD) in figure 3.

Figure 3:   Menu of senses for *sharp* in the OMD
          https://www.macmillandictionary.com/dictionary/british/sharp

The menu in figure 3 is a mere summary of the senses that will be presented and this form of assistance is useful especially in the case of articles of words with multiple senses. By looking at this menu the user who is interested in sense 5, for example, can save time by skipping the subcomments on semantics

in which senses 1–4 are presented and go directly to the subcomment on semantics containing sense 5.

A second approach to boxing different senses is to box sense headings separately as in *Cambridge Dictionary* (CD) in figure 4.

Figure 4:   Data boxes for *dull* in CD
https://dictionary.cambridge.org/dictionary/english/dull

In figure 4 in comparison to figure 2 the headings are not numbered nor given together but separately boxed at the start of each subcomment on semantics. The boxed information in figure 4 can be regarded as navigational devices, i.e. guide words. Taken at face value, words such as TEDIUS, UNINTERESTING, CLOUDY and STUPID in figure 5, are comparable to the boxed sections in figure 4 but words given in capital as well as lower case letters in figure 5 could be viewed as definitions.

Figure 5:   *Dull* in MW
https://www.merriam-webster.com/dictionary/dull
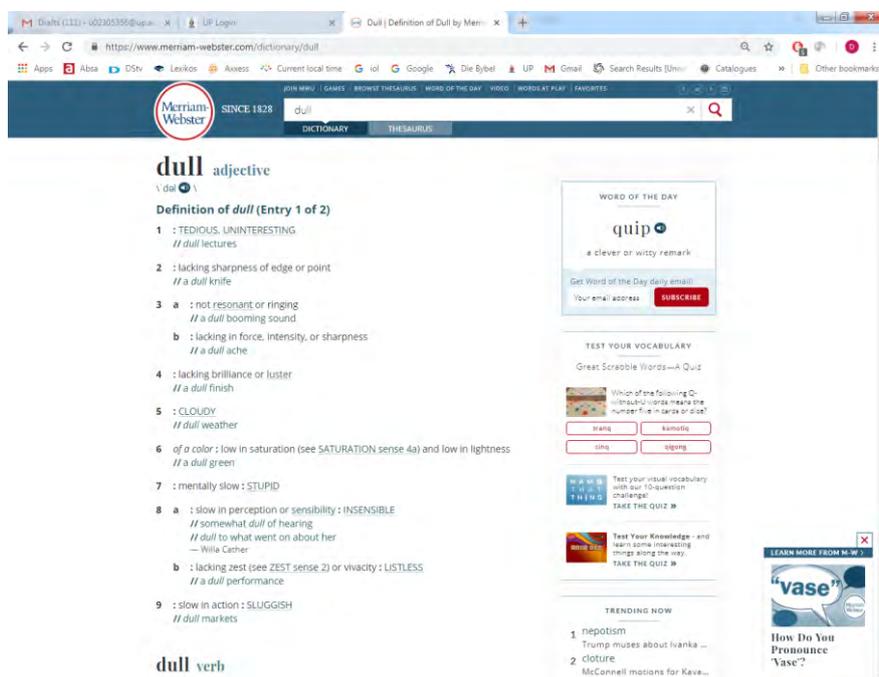
A third type of proposed low level data boxes is collocation boxes. The aim is to provide or bring together collocations of the lemma or derivatives and phrases in which it occurs in a data box as in figure 6.
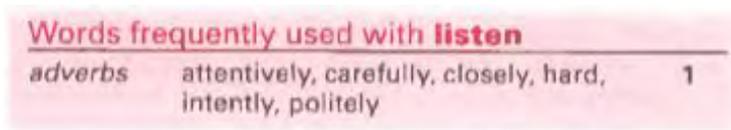
Figure 6:   Data box at *listen* in MED

The data box in figure 6 is useful to the reader looking up the word *listen* since it provides the typical collocations *attentively*, *carefully*, etc. in a box with the default treatment of *listen*.

Once again, the possible gain is on visibility — these collocations could be unboxed and presented, e.g. at the end of the article or in a search zone allocated to collocations.

A fourth type can be regarded as mere note boxes as appropriately labelled as such in (OALD). Consider figure 7 as a typical example for the data boxes linked to *be²* in OALD. The entry brings together the different forms of the present and past tenses of the verb *be* under the heading "NOTE".

Figure 7:   Data boxes for *be* in OALD

Figure 7 indicates what could be called *note* data boxes. The presentation starts with a horizontal line, followed by a white-on-black background capitalised label "note" and the present and past tenses boxed with full borders inside the note box amidst additional text. The note box as a whole does not have vertical lines on the left and right sides but is concluded by another horizontal line.

The *Oxford Dictionary of English* (ODE) uses data boxes for phrases and derivatives as in figure 8.
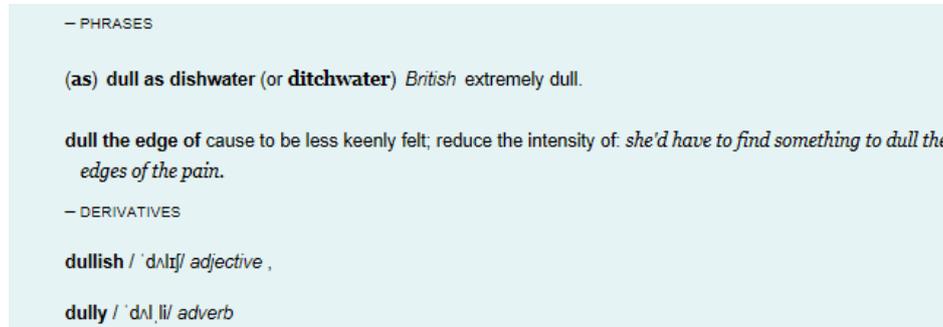
Figure 8:   Data box for phrases and derivatives of *dull* in ODE
http://www.oxfordreference.com.uplib.idm.oclc.org/view/10.1093/acref/9780
199571123.001.0001/m_en_gb0248630?rskey=7sNzVf&result=27211

MW uses data boxes for navigation as in figure 9.

Figure 9:   Data box for navigation of *dull* (column 1) and address of the Syno-
nyms & Antonyms hyperlink (column 2) in MW
https://www.merriam-webster.com/dictionary/dull

The guidance given in bottom tier boxes can also be conveyed by other means
that are employed in various dictionaries. These means, which will not be dis-
cussed here, include shortcuts, as found in the OALD, signposts, as used in the
LDOCE, and guide words, as presented in the *Cambridge Advanced Learner's Dic-
tionary* (CALD).

2.2   A middle tier of data boxes

This type of data box gives salient information that is not conveyed by items in

the default search zones of the articles of a specific dictionary such as items giving the paraphrase of meaning, translation equivalent paradigms and examples of usage. Typical boxes deal with guidance in terms of grammar, pronunciation, sense distinction, contrasting related words, restrictions on the range of application, register, spelling, pronunciation, etc.

2.2.1   Data boxes used to contrast related words

Typical of this type of data box is contrasting two or more words or different senses of the same word in variations of the language as in figure 10.

Figure 10:   Data box at *underground* in OALD
                 http://oald8.oxfordlearnersdictionaries.com/dictionary/underground_3

In figure 10 a data box linked to the first sense of underground nicely contrasts *underground*, *subway*, *metro* and *tube* in a very economical way. The same data box content is presented in the online *Oxford Learners Dictionaries* (https://www.oxfordlearnersdictionaries.com) but under the clickable menu item "+ British/ American underground / subway / metro / tube" as in figure 11.

underground *noun*

🔊 /ˈʌndəɡraʊnd/

🔊 /ˈʌndərɡraʊnd/

1 ⭐ *(often* **the Underground***) (British English)*
   (North American English **subway***)*

   [singular] an underground railway system in a city

   ▪ *underground stations*

   ▪ *the London Underground*

   ▪ *I always travel by underground.*

   SEE ALSO **London Underground**

   ─ British/American

   underground / subway / metro / tube

   ▪ A city's underground railway system is usually called the **underground** (often **the Underground**) in *British English* and the **subway** in *North American English*. Speakers of *British English* also use **subway** for systems in American cities and **metro** for systems in other European countries. **The Metro** is the name for the systems in Paris and Washington, D.C. London's system is often called **the Tube**.

Figure 11:   Data box at *underground* in OALD
https://www.oxfordlearnersdictionaries.com/definition/english/underground_3

This databox is repeated at *metro, tube* and *subway*.

In figure 12 the data box for *pavement* contrasts British versus American English.

pavement /ˈpeɪvmənt/ *noun* ★★
1 [C] *Br E* a path with a hard surface beside a road. *Am E* **sidewalk**
2 [U] *Am E* the surface of a road

Differences between British and American English: pavement

In the UK, a **pavement** is the hard raised level surface at the side of a road that people can walk on: *I set it down on the pavement by the door of the shop.* American speakers call this a **sidewalk**. In the US, **pavement** means the hard surface of a road: *Cars were skidding on the pavement.*

Figure 12:   Data boxes at *pavement* in MED

Consider also an isiZulu example for *kungathi* versus *sengathi* in figure 13.

kungathi ° *copulative cl. 17* ☻ SENGATHI'
■ it's as if; it seems as if; it looks as if; it
appears as if ● **Kungathi** bakhathele
kakhulu manje. *It looks as if they are very
tired now.*

The form "kungathi" can also be equivalent to
the conjunction "sengathi" (as if).
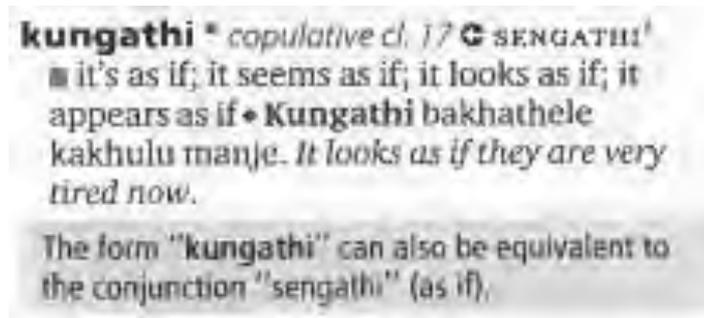
Figure 13:   Data box at *kungathi* in OZSD

Once again, it has to be stated that lexicographers are under no obligation to
provide data boxes for contrasting words — they could opt for alternative
strategies or even not to contrast the words at all. *Pharos Major Dictionary* (PMD)
treated *percentage point* as a sublemma in an article niche attached to the article
of the main lemma *percentage* and provides a data box at the end of the article
niche as in figure 14. The data box gives valuable additional information on
percentage point and contrasts *percentage* and *percentage point* very well. In the
presentation and treatment of *percentage point* in this case the compilers opted
for a single subarticle where the default data type, i.e. a translation equivalent.
is given but it is supplemented by an article-external data box. MED, however,
takes a different approach by lemmatising and treating *percentage* and *percent-
age point* in two separate main articles without a data box or any effort to relate
them as in figure 15.

percen'tage, persentasie; persentsgewys(e); *on a* ~
    BASIS, persentsgewys(e); *BY* ~*s,* persentsge●
    wys(e); ~ **point,** persentasiepunt; ~ **error,**
    persentasiefout.

> **Percentage point** refers to a (decimal) part of a
> percentage, e.g. if inflation rises by one percen●
> tage point (say from 10 to 11%), the actual per●
> centage rise is 10%, while the rise in percentage
> points is 1%.

Figure 14:   *Percentage* and *percentage point* in PMD
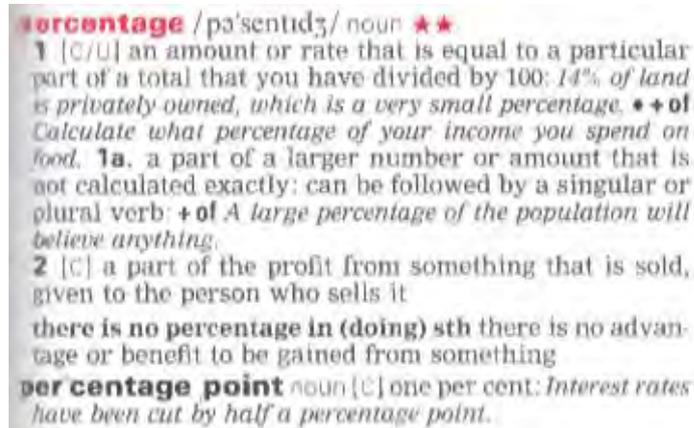
Figure 15:   *Percentage* and *percentage point* in MED

2.2.2   Data boxes focused on application range or restrictions

This type of data boxes guides the user in terms of the contexts in which a word can be used as well as instances where the use of such a word would be inappropriate. Consider figure 16.

Figure 16:   Data box for *maritime* in RWD

In figure 16 the data box for *maritime* explains the meaning of *maritime* as 'adjacent to the sea' but that it should not be used to refer to a house at the seaside.

2.2.3   Data boxes providing grammar information

Data boxes giving guidance to correct grammatical use cover a variety of

aspects such as the use of singular versus plural forms, tense forms of verbs, translations, abbreviated and irregular forms, etc. Consider figures 17 and 18:
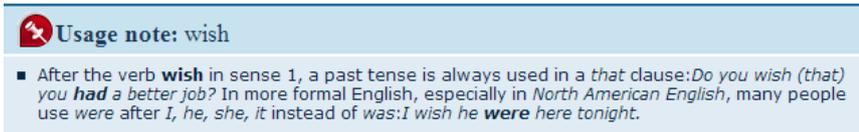
> ❌ **Usage note:** wish
>
> ■ After the verb **wish** in sense 1, a past tense is always used in a *that* clause:*Do you wish (that) you **had** a better job?* In more formal English, especially in *North American English*, many people use *were* after *I, he, she, it* instead of *was:I wish he* **were** *here tonight.*

Figure 17:   Data box at *wish* in OALD
                http://oald8.oxfordlearnersdictionaries.com/dictionary/wish_1

The use of *wish* in figure 17 is restricted on grammatical grounds, i.e. in terms of tense and nature of the following verb.

The noun **lefase** is sometimes found without the prefix **le-**; thus as **fase**.

Figure 18:   Data box for *lefase* in ONSD

In figure 18 the data box for *lefase* indicates its use without the prefix. Consider also the data boxes for *student/studente* and *neither* in figure 19. It indicates that the form *studente-* is required for use as the first part of a compound and that *neither* should be followed by a singular noun, etc.

Figure 19:   Data box for *student, studente* in RWD and *neither* in OALD

Finally the data boxes in figures 20 and 21 deal with the important issues, i.e. (a) that *the*, *a* and *an* do not have translation equivalents in isiZulu; (b) in certain cases subject concords are not translated [*di*[1] and *le*[3]] and (c) providing grammatical information on tense form of an irregular verb [*-shongo*].

Figure 20:   Data boxes for *the* and *shongo* in OZSD

Figure 21:   Data boxes for *di* and *le* in ONSD

2.2.4   Data boxes for pronunciation guidance

Pronunciation guidance is usually given in the default treatment of the lemma by means of descriptions, respelling or phonetic symbols, but specific pronunciation issues such as pronunciation comparison with other words can be given in data boxes. In figure 22 the "o" in *brons* is described in terms of the basic characteristics of "short" and "long".

Figure 22:   Data box for *brons* in RWD

In figure 23 guidance in pronunciation of words ending in *-et*, presented in the partial article stretch between the articles of *et al.* and *etc.* in *Cambridge International Dictionary of English* (CIDE), is given by means of phonetic transcriptions and stress on syllables.

PRONUNCIATION OF WORDS ENDING IN *-ET*

In most words ending in 'et' the final 't' is pronounced: *bullet* /ˈbʊl·ət/ *supermarket* /£ˈsuː·pə·mɑː·kɪt/ /$ˈsuː·pɚ·mɑːr·kɪt/. However, some words ending in 'et' are borrowed from French: in these words, 'et' is pronounced /eɪ/.

ballet	crochet
beret	croquet
bidet	duvet
bouquet	gourmet
buffet	parquet
cabaret	ricochet
cachet	sorbet
chalet	

In British English, the first syllable of these words is stressed; in American English, the final syllable with 'et' is stressed /£ˈbæl·eɪ/ /$bæˈleɪ/ • /£ˈber·eɪ/ /$bəˈreɪ/
Bouquet does not follow this rule: some British speakers stress the second syllable.

There are a few other words with a silent final t:

debut /£ˈdeɪ.bjuː/ /$deɪˈbjuː/
rapport /£ræˈpɔːʳ/ /$ræˈpɔːr/

Figure 23:   Data box for pronunciation of words at *et al.* in CIDE

2.2.5   Data boxes indicating register

Data boxes on register deal with issues such as *formal/informal* and *written versus spoken language*.
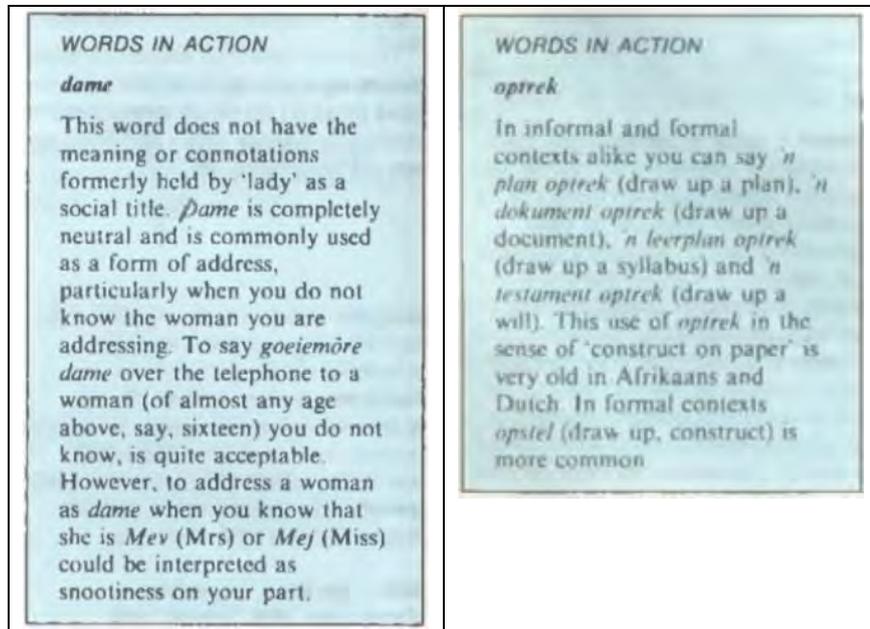
Figure 24:   Data boxes for *dame* and *optrek* in RWD

In figure 24 the data box reflects on change of meaning and connotations of the Afrikaans word *dame* compared to its English equivalent *lady*, and the contexts in which the use of this word is acceptable or not. The data box for *optrek* gives guidance on formal versus informal use as well as mentioning antiquation in certain senses.
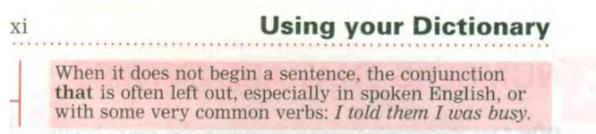
Figure 25:   Data box at *that* in OALD

In figure 25, among other aspects, guidance is given on the use or omission of *that* in spoken language.

2.2.6   Data boxes dealing with spelling

This type of data boxes mainly deals with spelling variants, capitalization and word divisions.
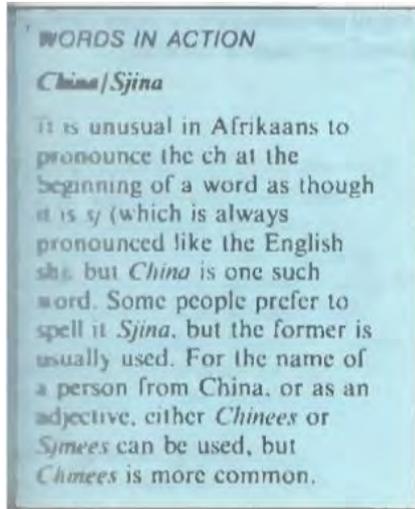
Figure 26:   Data box for *China*/*Sjina* in RWD

In figure 26 the data box indicates that both spelling variants, i.e. *China* and *Sjina,* are acceptable in Afrikaans.
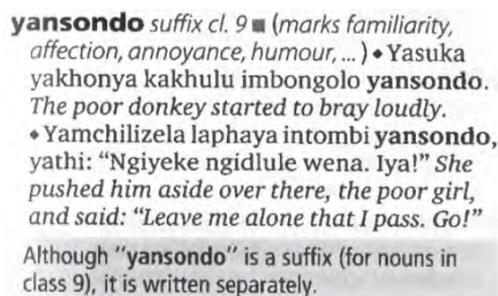
Figure 27:   Data box for *yansondo* in OZSD

In figure 27 the data box deals with word division, i.e. that this nominal suffix is written separately.

2.2.7   Data boxes indicating syntactic restrictions.

This type of data boxes mainly gives guidance on syntactic positions of words in sentences. Consider the following examples that are only relevant for Sepedi and isiZulu respectively in figures 28 and 29.

Figure 28:   Data boxes for *afa* and *afaeya* and *kilego* in ONSD

In Sepedi, the question particles *afa* and *afaeya* in contrast to the question parti-
cle *na* are restricted to the sentence-initial position. The auxiliary verb stem
-kilego, which is in the relative mood is followed by the consecutive.

Consider also the data boxes given for the isiZulu words *ngabe* and *lena* in
figure 29. These boxes indicate that *na* cannot be used sentence-initially but that
it is permissible for *ngabe* and that the demonstrative pronoun *lena* has to be
used post-nominally.

Figure 29:   Data boxes for *ngabe* and *lena* in OZSD

These are also good examples of a language specific issue for an African lan-
guage not applicable to other members of the language family as mentioned
above.

2.2.8   Data boxes dealing with obsolete, archaic and antiquating words

This type of data box has its finger on the pulse of a language in terms of lan-
guage change. We regard "obsolete" and "archaic" in terms of MED as "no
longer used" and "antiquating" as becoming obsolete, cf. figure 30.
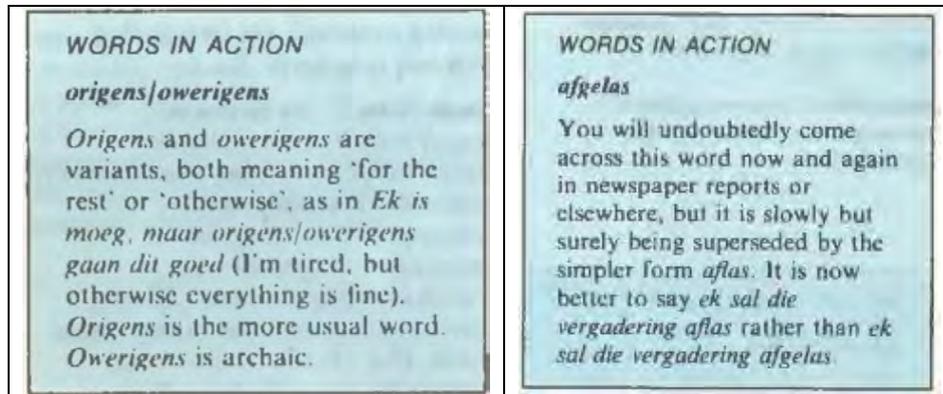
Figure 30:   Data boxes for *origens/owerigens* and *afgelas* in RWD

In figure 30 it is indicated that although *origens* and *owerigens* have the same meaning, *owerigens* became archaic. The same holds true for *afgelas* in the sense of the intended cancellation of, e.g. a meeting, which is antiquating in favour of *aflas*.

2.3   A top tier of data boxes

The proposed top tier of data boxes is distinguished for providing users with indispensable salient data of a serious nature regarding warnings, taboos and even illegal words. Even inside the category of top tier, a hierarchy can be distinguished ranging from mere recommendation in the sense of 'often considered insulting' to 'avoid using this word' to 'absolutely forbidden to use', i.e. of which the use is a criminal offence and punishable by law.

In figure 31 the data box at *umfazi* in OZSD is an example of a mere recommendation, i.e. where a better option is suggested.
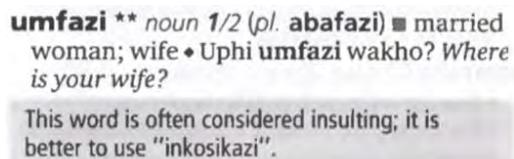

Figure 31:   Data box for *umfazi* in OZSD

The data boxes in figures 32, 33 and 34 suggest a stronger condition, i.e. avoidance of the words *crazy*, *old* and *deaf mute* when referring to a person.
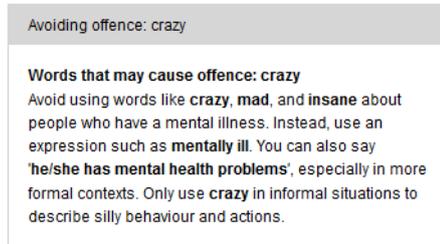
Figure 32:   Data box at *crazy* in the Macmillan Dictionary Online
http://www.macmillandictionary.com/dictionary/british/crazy#crazy_4
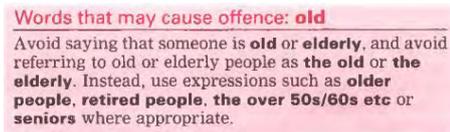
**Figure 33:**   Data box for *old* in MED
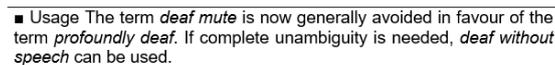
Figure 34:   Data box for *deaf mute* in *The Concise Oxford Dictionary* (COD)

In the (South) African context a number of words, mostly words insulting black people, exist that are considered to be so offensive that it is illegal even to say or write these words. Aliases have to be used if reference to such words are absolutely necessary e.g. in media reports or the judicial system e.g. the *k-word, n-word, h-word, m-word* etc.

In 1994 the Bureau of the Woordeboek van die Afrikaanse Taal (WAT) made a sincere attempt to address this issue by organising an international conference on the handling of insulting and sensitive lexical items in order to formulate a policy on the handling of such lexical items in the WAT. Harteveld and Van Niekerk (1995: 233) report on the outcome of this conference and state that the point of departure of the WAT was to fulfil its ideal of comprehensiveness but also to follow a policy of sensitive handling of lexical items.

> Die Buro van die WAT wil in sy strewe na omvattendheid nie aandadig wees aan die vestiging of bestendiging van rassistiese leksikale items deur die opname daarvan in die WAT nie, maar hy het wel 'n verantwoordelikheid om gebruikers te waarsku teen die rassistiese aard van sekere leksikale items. Dit kan hy slegs doen as hy hierdie leksikale items identifiseer en op een of ander wyse onder die aandag van die gebruiker bring. (Harteveld and Van Niekerk 1995: 235)

(The Bureau of the WAT, in its pursuit of comprehensiveness, does not want to be complicit in the establishment or perpetuation of racist lexical items by including them in the WAT, but it does have a responsibility to warn users against the racist nature of certain lexical. items. He can only do this and if he identifies these lexical items and somehow brings them to the attention of the user.)

The dilemma of lexicographers is clear — on the one hand they do not want to contribute to the use of offensive lexical items by including them in the dictionary but on the other hand feel a strong responsibility to reflect the lexicon of the specific language and, especially, to warn their users against the use of offensive terms.

3.   A summary of data box types in RWD, ONSD and OZSD

The final section of this article reflects a survey that was made of all data boxes in the Afrikaans to English side of RWD as well as the Sepedi to English and English to Sepedi side in ONSD and the isiZulu to English and English to isiZulu sides of OZSD.

In the Afrikaans to English side of RWD no less than 2,000 data boxes were provided as broken down in descending order in terms of type and given per alphabetical stretch in table 1.

Table 1:   Data boxes in RWD (Afrikaans to English side)

alph stretch	communica-tion/applica-tion range or restrictions	contrast related words	grammar info	syntactic restrictions	register and written/ spoken	contrast BE, AE + angli-cisms + origin	obsolete, antiquating/ freq.	spelling	pronun ciation
A	41	40	20	33	35	20	1	2	4
B	36	33	21	18	13	19	3	7	3
C			2		1	3		5	3
D	17	22	7	13	8	13	4	7	3
E	9	7	7	6	4	2	4	2	
F	5	4	3	1	3	1	2		
G	16	17	19	17	9	3	2	5	4
H	16	15	17	12	13	6	2	4	5
I	5	5	10	6	4	1	4	4	2
J	2	3	14	1	4	1	1	2	
K	23	10	24	28	8	10	20	3	7
L	16	2	16	7	6	4	9	2	
M	28	37	17	2	10	5	5	6	1
N	9	4	6	5	9	2	5	2	1

O	43	17	10	24	14	11	10	5	5
P	13	5	13	7	14	11	10	7	2
Q		1			1				
R	13	11	5	9	5	6	9	6	5
S	46	25	28	24	26	16	21	12	3
T	30	7	5	5	8	5	4	7	1
U	10	8	4	5	4	2	3	3	
V	54	21	18	17	26	14	10	10	2
W	29	16	17	4	11	4	6	11	1
X									
Y	1	1		1			1	1	1
Z								1	1
	462	**311**	**283**	**245**	**236**	**159**	**136**	**114**	**54**
Total									2000
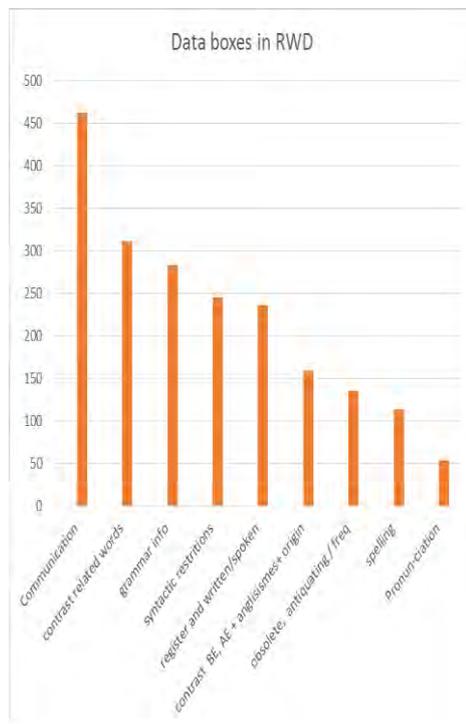
Figure 35:   Data boxes in RWD (Afrikaans–English side)

From table 1 and figure 35 it is clear that the top five types of data boxes deal with issues related to range of application, restrictions, contrast, grammar, syntactic restriction and register. The 2,000 data boxes presented in 639 pages give an average of approximately 3 boxes per page.

Consider the content summary of data boxes in the alphabetical stretches for M in RWD (Gouws and Prinsloo 2010: 507) in table 2 with rank comparisons of these categories between the two sides in table 3.

Table 2: Types of data boxes in common: *M* in Afrikaans–English and *M* in English–Afrikaans in RWD

M **Afr–Eng** *RWD*		*M* **Eng–Afr** *RWD*	
Contrast related words	37	Communication/application range or restrictions	38
Communication/application range or restrictions	28	Contrast related words	23
Grammar data	17	Contrast BE, AE	12
Register and written/spoken	10	Register	11
Spelling	6	Syntactic restrictions	8
Contrast BE, AE + Anglicism + origin	5	Pronunciation	7
Obsolete, antiquating / freq.	5	Grammar data	7
Syntactic restrictions	2	Spelling	3
Pronunciation	1	Obsolete, antiquating	3

Table 3: Rank comparison of *M* in Afrikaans–English and *M* in English–Afrikaans in RWD

M **Afr–Eng** *RWD*		*M* **Eng–Afr** *RWD*	
	Rank in Afr–Eng	*Rank in Eng–Afr*	*Rank differences*
Contrast related words	1	3	2
Communication/application range or restrictions	2	1	1
Grammar data	3	7	4
Register and written/spoken	4	4	0
Spelling	5	8	3
Contrast BE, AE + Anglicism + origin	6	3	3
Obsolete, antiquating / freq.	7	9	2
Syntactic restrictions	8	5	3
Pronunciation	9	6	3
		Average rank difference:	**2.3**

From the rank comparisons in table 3 it is clear that the average rank difference is very small indicating similarity in the types and contents of data boxes in the Afrikaans–English and English–Afrikaans sides.

The types of data boxes used in the Sepedi to English and English to Sepedi sides of ONSD are given in table 4 and graphically illustrated in figure 36. The data types indicated in boldface in table 4 indicate the types of data boxes that occur on both sides of the dictionary.

Table 4:   Data boxes in ONSD

Sepedi–English ONSD		English–Sepedi ONSD	
Content	**Number**	**Content**	**Number**
demonstratives	37	complex words	20
aux./cop. verb/conjunctive/	34	**contrast & don't confuse**	16
used as different part of speech	30	info pronunciation	11
composition (consists of x+y)	19	right and wrong	6
subject concords	15	**range of application**	4
shortened forms	15	articles not in Sepedi	3
range of application	13	combined with other	3
state of existence	12	**different spelling**	1
spelling guidance	6	formal/informal	1
question particles	5	cross-reference box	1
unique function	3	abbreviate	1
unusual spelling of	3	singular/plural	1
past tense marker	1		
irregular form	1		
writing versus pronunciation	1		
contrast	1		
offensive	1		
combined with other	1		
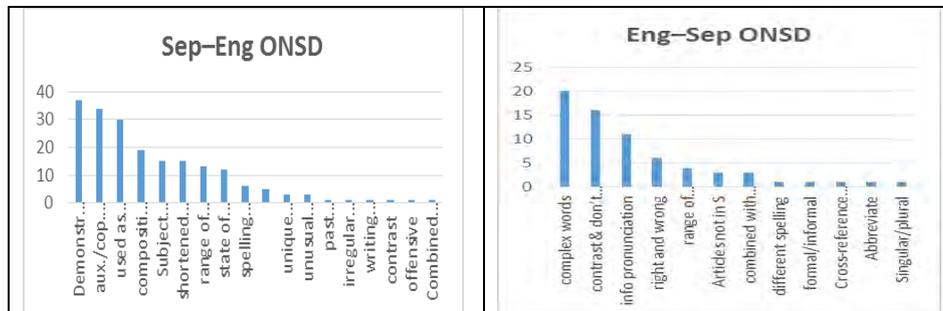
Figure 36:   Data boxes in ONSD

Most of the data boxes in the Sepedi to English side give guidance on the nature and use of demonstratives while most data boxes on the English to Sepedi side deal with complex words.

Data boxes giving guidance on equivalents and ways to express concepts top the list of data box contents in the English to isiZulu side and data boxes dealing with grammatical issues pertaining to syntax, tense and extended or shortened forms being the most frequent in the isiZulu to English side, cf. table 5 and figure 37.

Table 5:   Data boxes in OZSD

English–isiZulu		isiZulu–English	
Content	Number	Content	Number
different equivalent / concept expression	40	**grammar (esp. syntax, tense, extended/shortened forms)**	39
pronunciation	27	demonstratives & time/space relations	35
do not confuse & contrasting words	21	**meaning, contrast & range of application**	20
grammar	8	nouns, etc. used as adverbs	13
register	6	**spelling and word division**	12
sing/plural	5	compare with English & translation	8
spelling	3	**cross-reference**	1
cross-reference	2	offensive	1
collocations	1		
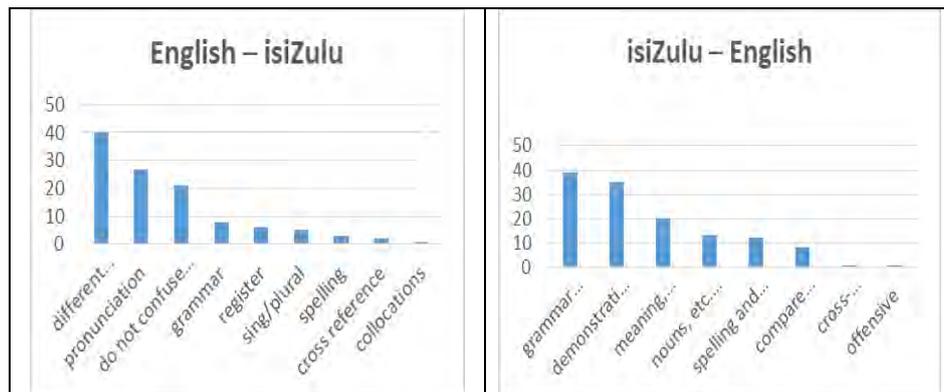
Figure 37:   Data boxes in OZSD

4.   Conclusion

In Part 1 (this volume) the focus was on data boxes as text constituents. This article focused on the types and contents of data boxes and in Part 3 guidance will be offered for prospective compilers on data boxes of the future. In Part 2 it was emphasized that no structural planning of data boxes nor specific user-guidance on the nature and use of data boxes or distinction between different types of data boxes was observed in the dictionaries studied. Data boxes are presented in a haphazard way without any clear treatment convention and conformity. What lies beyond doubt, however, is that all the sources quoted

above express a need for a lexicographic strategy to help users to avoid common mistakes, get additional information, learn more about the word and its origin, etc. The focus was on the analysis of data boxes in existing dictionaries to determine the nature of data presented in boxes and a three-part hierarchy was suggested. The first type was labelled as the mere bringing together and highlighting of aspects such as menus for the different senses of the word and lists of typical collocations. The second type, a much larger and more diverse category dealt with data boxes providing salient information which falls outside the default lexicographic treatment devices such as paraphrase of meaning, translation equivalent paradigms and examples of use. The final category represents the top tier in the proposed hierarchy namely data boxes for restricted words in terms of warnings and alerts to their use or avoidance.

Acknowledgement

References

Dictionaries

CALD = McIntosh, Colin (Ed.). 2013. *Cambridge Advanced Learner's Dictionary.* 4th edition. Cambridge: Cambridge University Press.

CD = *Cambridge Dictionary.* https://dictionary.cambridge.org/dictionary/english/. [Accessed 1 August 2020.]

CIDE = Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English.* Cambridge: Cambridge University Press.

COD = Thompson, Della (Ed.). 1995. *The Concise Oxford Dictionary.* 9th Edition. Oxford: Clarendon Press.

LDOCE = Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English.* Harlow: Longman.

MED = Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.

MW = *Merriam Webster.* https://www.merriam-webster.com. [Accessed 1 August 2020.]

OALD = Turnbull, J. (Ed.). 2010. *Oxford Advanced Learner's Dictionary of Current English*. 8th edition. Oxford: Oxford University Press. [Accessed 24 December 2013.]

OALDC = Crowther, Jonathan (Ed.). 1995. *Oxford Advanced Learner's Dictionary of Current English.* 5th edition. Oxford: Oxford University Press.

ODE = *Oxford Dictionary of English.* http://www.oxfordreference.com.

OMD = *Macmillan Dictionary — British English Edition.* https://www.macmillandictionary.com/dictionary/british/. [Accessed 2 August 2020.]

ONSD = De Schryver, G.-M. (Ed.). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford.* Cape Town: Oxford University Press Southern Africa.

OZSD = De Schryver, G.-M. (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford*. Cape Town: Oxford University Press Southern Africa.

PMD = *Pharos Major Dictionary*. Eksteen, L.C. (Ed.). 1997. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans.* 14th expanded edition. Cape Town: Pharos.

RWD = Grobbelaar, P. et al. (Eds.). 1987. *Reader's Digest Afrikaans–Engelse Woordeboek / English–Afrikaans Dictionary*. Cape Town: The Reader's Digest Association, South Africa (Pty) Ltd.

Other references

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Gouws, R.H. and D.J. Prinsloo. 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010:* 501-511. Leeuwarden: Fryske Akademy.

Harteveld, P. and A.E. van Niekerk. 1995. Beleid vir die hantering van beledigende en sensitiewe leksikale items in die *Woordeboek van die Afrikaanse Taal* (as formulated by the authors). *Lexikos* 5: 232-248.

Lexicographic Data Boxes
Part 3:
Aspects of Data Boxes in
Bilingual Dictionaries and
a Perspective on Current and
Future Data Boxes[*]

Rufus H. Gouws, *Department of Afrikaans and Dutch,*
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)
and
D.J. Prinsloo, *Department of African Languages, University of Pretoria,*
Pretoria, South Africa (danie.prinsloo@up.ac.za)

Abstract: This article, the third in a series of three on lexicographic data boxes, firstly focuses on a number of aspects of data boxes in bilingual dictionaries with the emphasis on different approaches in bilingual dictionaries with an African language as one of the members of the treated language pair. It is not possible to provide a comprehensive discussion within the limitations of an article. Then the discussion proceeds by looking at some new ways of using data boxes in online dictionaries. It is shown that the possibilities of the new medium allow lexicographers to employ data boxes in both traditional and non-traditional ways. It is argued that data boxes are expected to fulfil a variety of purposes ranging from navigational information and the provision of salient information to giving access to relevant data in dictionary-internal and dictionary-external sources. Lexicographers of online dictionaries have introduced new ways of using data boxes that have not yet been fully discussed in metalexicographic literature. This article gives an identification and a brief discussion of some of these innovative uses of data boxes. It stresses the potential that the online environment offers lexicography. Practical and theoretical lexicographers need to be aware of these possibilities and challenges. By embarking on a more comprehensive use of data boxes dictionaries can become even better containers of knowledge and can serve their users in an optimal way.

Keywords: DICTIONARIES, DATA BOXES, POP-UP BOXES, HYPERLINKING, AFRICAN LANGUAGES, SEPEDI, ISIZULU, SEARCH DOMAIN, SEARCH UNIVERSE, DATA DISTRIBUTION

Opsomming: **Leksikografiese datakassies. Deel 3. Aspekte van datakassies in tweetalige woordeboek en 'n perspektief op huidige en toekomstige data-**

---

[*] This is the third in a series of three articles dealing with various aspects of lexicographic data boxes.

kassies. Hierdie artikel, die derde in 'n reeks van drie oor leksikografiese datakassies, fokus eerstens op aspekte van datakassies in tweetalige woordeboeke met die klem op verskillende benaderings in tweetalige woordeboeke met 'n Afrikataal as een van die lede van die behandelde taalpaar. Daarna gaan die bespreking voort deur te kyk na 'n paar nuwe maniere om datakassies in aanlyn woordeboeke te gebruik. Daar word aangetoon dat die moontlikhede wat die nuwe medium bied, leksikograwe in staat stel om datakassies op sowel tradisionele as nie-tradisionele maniere te gebruik. Daar word aangevoer dat die datakassies gebruik kan word om 'n verskeidenheid doel-eindes te bereik, wat wissel van navigasie-inligting en die verskaffing van belangrike inligting tot toegang tot relevante data in woordeboek-interne en woordeboek-eksterne bronne. Leksikograwe van aanlyn woordeboeke het nuwe maniere bekendgestel om datakassies te gebruik wat nog nie volledig in die metaleksikografiese literatuur bespreek is nie. Hierdie artikel gee 'n identifisering en 'n kort bespreking van sommige van hierdie innoverende gebruike van datakassies. Dit beklem-toon die potensiaal wat die aanlynomgewing aan die leksikografie bied. Leksikograwe moet bewus wees van hierdie moontlikhede en uitdagings. Deur met 'n meer omvattende gebruik van datakas-sies te begin, kan woordeboeke selfs beter kennishouers word en hul gebruikers op 'n optimale manier dien.

Sleutelwoorde: WOORDEBOEKE, DATAKASSIES, OPSPRINGKASSIES, HIPERSKAKELS, AFRIKATALE, SEPEDI, ISIZULU, SOEKDOMEIN, SOEKUNIVERSUM, DATAVERSPREIDING

1.   Introduction

The fact that data boxes are not used at all in many dictionaries and that they are often almost randomly used merely to bring together or highlight informa-tion could create the impression that data boxes have an insignificant role to play in dictionaries and should therefore only belong to the periphery of metalexicographic discussions. In this article we wish to argue to the contrary, i.e. that data boxes are important and even essential but underutilized lexico-graphic components which should be used to fulfil specific needs. The user should be guided with regard to salient information which cannot typically be catered for by standard dictionary conventions and items such as those giving a paraphrase of meaning, translation equivalents and examples of use. We wish to make it clear that treatment in data boxes is not in competition with the default treatment in the article of a specific lemma. They supplement each other, default treatment is the first objective then consideration of a data box if the compiler deems it necessary to give further guidance on salient informa-tion. In Part 1 Gouws and Prinsloo (2021) and Part 2 Prinsloo and Gouws (2021) (this volume), lexicographic data boxes as text constituents in dictionaries and types and contents of data boxes were discussed with the purpose of setting the scene for the discussion of the future of data boxes and data boxes of the future in this article.

Moreover, as seen in section 3, data boxes did not lose their relevance in the transition from paper to electronic dictionaries but electronic dictionaries often employ alternative strategies for the presentation of data boxes enabled by the digital era.

Dictionaries have a genuine purpose, cf. Wiegand (1998: 299). This also applies to the different components of dictionaries, including the data boxes. Data boxes of the future should focus on what we believe is the genuine purpose, cf. Wiegand (1998: 299), of data boxes, i.e. guidance on salient data not sufficiently emphasised in the default lexicographic presentation. These include for example the contrasting of different words, aspects of the range of application, antiquation, taboos etc. However, the possibilities offered by the online environment and the innovative and dynamic options regarding the structure of dictionary articles should lead lexicographers to use data boxes in ways that include but also go beyond the mere representation of salient data. Where the first part of this article focuses on various aspects of data boxes presenting salient data, the second section moves towards new uses of data boxes. In the first section the discussion will be directed at some issues in bilingual dictionaries in which an African language is a member of the treated language pair. The second section will primarily be directed at online dictionaries in general but will also be relevant to future African language dictionaries.

2.   Data boxes in bilingual dictionaries — African languages as a case in point

2.1   Different approaches to data boxes

Dictionaries for the African languages could firstly cater for the inclusion of data boxes dealing with issues not restricted to the given language pair. Secondly, they could include data boxes specific to the language family they belong to and finally data boxes dealing with unique features of individual members of the language family. With regard to these three issues data boxes can play a significant role in making the user aware of salient data. What is presented in this paragraph is a selection of a number of lemmas which should be considered for the provision of data boxes in addition to the standard treatment given in the dictionary article. The selected issues pertain to different aspects of morphology, syntax and semantics such as (a) demonstratives, (b) multiple recurring phrases as translation equivalents, (c) reference to men and women versus addressing them, (d) different constructions used for English adjectives and (e) equivalent relations. It is, however, not possible to present a comprehensive or systematic account of the full scope of required data boxes within the limitations of an article.

The African language Sepedi (a Sotho language) will be taken as example language in the following discussion with occasional reference to isiZulu (a Nguni language) both belonging to the Bantu[1] Language Family. To the knowledge of the authors the only Sepedi and isiZulu dictionaries using data boxes are the Oxford school dictionaries for Sepedi and isiZulu (henceforth ONSD and OZSD respectively).

As far as the first category is concerned it can simply be stated that data

boxes for African languages should give guidance on issues applicable to all languages such as contrasting related words, range of application, cultural considerations, etc. Secondly, attention should be given to typical characteristics of the language family such as verbal moods, nominal classes, kinship terminology, etc. Finally, data boxes should be included guiding the user on issues characteristic of the specific African language such as guidance on pronunciation, syntax, semantics, word division, etc. These issues are discussed in detail in Part 1 Gouws and Prinsloo (2021) and Part 2 Prinsloo and Gouws (2021). Consider figure 1 as an example dealing with contrast and range of application.
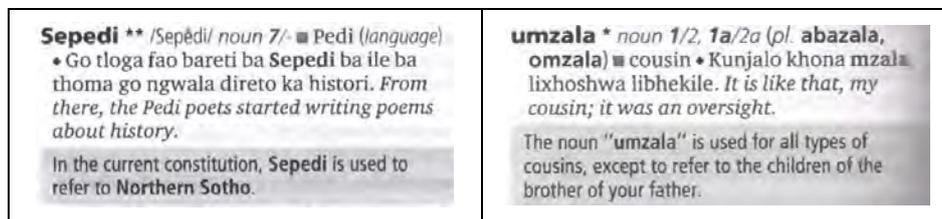
Figure 1:   Data boxes for *Sepedi* and *umzala* in ONSD and OZSD — contrast and range respectively

The data box at *Sepedi* contrasts the names *Sepedi* versus *Northern Sotho* and informs the user that these terms refer to the same language. There is much controversy around the use of these names and what the relation or difference between these terms is, therefore guidance is required. The data box at *umzala* gives a precise indication of the range of application i.e. that it can be used to refer to cousins but not to the children of one's father's brother. A complicated system of kinship terminology exists for African languages, cf. Van Wyk and Haasbroek (1990) for Setswana and Prinsloo and Van Wyk (1992) for Sepedi. Data boxes can provide valuable guidance on kinship, e.g. on the range of application as in the case of *umzala* in figure 1 and contrasting kinship relations, e.g. relatives on father's versus mother's side, whether you address a specific relative versus speak about them. The shortage of space will always be a consideration in paper dictionaries and it is for the lexicographer to prioritise the type of information to be provided, e.g. treatment in the article by means of translation equivalents or providing a data box or both. So, for example, the compiler of OZSD and ONSD has valued the importance of text boxes at demonstratives as so important that he dedicated 40% of the page in figure 5 to text boxes.

As far as the second category is concerned, figure 2 illustrates a typical category which Sepedi and isiZulu (and probably all other members of the language family) have in common, i.e. guidance on grammatical data like nominal classes and their concords or that the English articles *a*, *an* and *the* are not translated/do not have equivalents.

bo¹ *** *subject concord cl. 14* ■ it ◆ Bogobe **bo** a fiša. > **Bo** a fiša. *The porridge is hot.* > *It is hot.*

Subject concords of noun classes, here **bo**, are not translated when they appear together with the subject in the sentence.

the *** *article* ■ - ◆ We crossed **the** river at **the** drift. *Sawela umfula ezibukweni.*

The English definite article *the*, as well as the indefinite articles *a* and *an*, have no direct translations in Zulu. In English, *the* is used before nouns.

Figure 2:   Data boxes at *bo* in ONSD and *the* in OZSD

As far as data boxes dealing with unique features of Sepedi and isiZulu are concerned, much guidance is required in respect of what Prinsloo (2017 and 2020) call complicated grammatical structures. Consider figures 3 and 4 as examples of unique features of prefixing and composition of specific words in Sepedi and isiZulu.

ngale ** *locative adverb* ↻ LE² ■ far away in that direction; far away on the other side ◆ Lowo mfana ubalekile waye washona **ngale**. *That boy ran away and eventually disappeared there, far away in that direction.*

The instrumental formative "nga-" can also be prefixed to the demonstrative pronoun "le", at which point the meaning of "**ngale**" is "with or about these/this".

Figure 3:   Data box at *ngale* in OZSD

In the treatment of *ngale* in figure 3 a cross-reference to the locative adverb *le²* is given. The data box, however, refers to the demonstrative pronoun *le¹* which is lemmatised and treated in its appropriate alphabetical position in OZSD.

nyakang *verb* ■ searching for what?; looking for what? ◆ Batho ba ba re ba **nyakang**? *What do these people say they are looking for?*

This form comes from the verb **nyaka** and the question word **eng**.

Figure 4:   Data box at *nyakang* in ONSD

For most African languages strong normative guidance is required since standardization is still in progress, cf. Gallardo (1980: 62). Such boxes could well be high on the list of typical data box content for these languages and OZSD and ONSD have done well in the provision of valuable information for the users in

data boxes. Consider the following six examples of data boxes for Sepedi dictionaries which could substantially enhance their value in respect of user guidance suggested as model entries for future Sepedi dictionaries.

2.2   Demonstratives

Demonstratives basically express *this* or *these* in relation to three relative distances within sight of the speaker, e.g. *monna yo* 'this man', *monna yoo* 'that man' and *monna yola* 'that man over there, yonder'. Linguists such as Louwrens (1991), Van Wyk et al. (1992), Lombard (1985) and Poulos and Louwrens (1994) distinguish three basic positions but differ in respect of the sub-positions into which demonstratives can be classified. Louwrens (1991) distinguishes between the different positions as follows:

Position 1(A) Speaker and the addressee are *close to one another*, while the object referred to is relatively *near them*
Position 1(B) Speaker and the addressee are *at a distance from each other*, while the object referred to is directly next to the speaker
Position 2(A) Speaker and the addressee are *relatively far apart*, while the object referred to is *nearer to the addressee*
Position 2(B) Refer to objects which are *very close or directly next to the addressee*
Position 3 Speaker and the addressee are *very close to one another*, while the object referred to is *far away* from them

Louwrens (1991: 106-108)

Consider table 1 as an extract from the table given in ONSD.

Table 1:   Extract from the table for demonstratives in ONSD

Cl.	Pref.	Pos. 1	Pos. 2	Pos. 3	Pos. 1a	Pos. 1b	Pos. 2b
1	mo-	yo	yoo	yola	yono	yokhwi	yowe
2	ba-	ba	bao	bale	bano	bakhwi	bawe
---	---	---	---	---	---	---	---
9	(N)-	ye	yeo	yela	yeno	yekhwi	yewe
10	di-	tše	tšeo	tšela	tšeno	tšekhwi	tšewe
---	---	---	---	---	---	---	---
14	bo-	bjo	bjoo	bjola	bjono	bjokhwi	bjowe
---	---	---	---	---	---	---	---

ONSD: S18 and S19

Two issues pertaining to demonstratives are relevant to the user i.e. firstly a complete table indicating all the demonstratives of the different positions and

classes and their basic meanings, i.e. indicating three distances, 'here', 'there' and 'there (yonder). Secondly indication of the exact semantic relations in respect of speaker and addressee is required. The lexicographer could, for example, give the full table, e.g. as in table 1 and the basic translations as the reference address e.g. in the back matter of a paper dictionary or as a clickable pop-up window in an electronic dictionary — see the next section. The purpose is a complete illustration of the different classes and positions of demonstratives and their meanings and translation equivalents. The meanings of the different positions as described by Louwrens (1991: 106-108) above could be presented as pop-up boxes for each demonstrative or as data boxes in the central text of a paper dictionary as has successfully been done in ONSD as in figure 5.

Figure 5:   Data boxes for demonstratives in Sepedi in ONSD

In figure 5 the compiler regarded the salient information given by data boxes for demonstratives as so important that they are provided for each demonstrative and not only as a separate section in e.g. the back matter of the dictionary. Such a decision remains the prerogative of the compiler in consideration of the skills level of the target user.

2.3   Multiple recurring phrases as translation equivalents

These are cases where an English word can be translated by means of a grammatical pattern determined by the different noun classes, — *every* as a typical case in figure 6.

Figure 6:   *Every* in ONSD

The duplication of the adjective construction, as in figure 6, i.e. Class 5 *le lengwe le le lengwe* (*le lengwe*: a certain/other (day) + *le*: and + *le lengwe*: another one) reflects a single instance of the recurring pattern for all other classes. This example is fine but it is important to inform the user by means of a data box that this can be done for all classes keeping in mind that the concords used have to match the nominal class to which the noun belongs. A data box as in figure 7 which gives examples from more noun classes is recommended at the article for *every*. Such a box should preferably have a cross-reference to the back matter where a full table with a description of the strategy, i.e. that the concept 'every' is expressed by means of the duplication of 'another' (e.g. *monna yo mongwe* 'another/certain man' *le* 'and' *yo mongwe* 'another one') is given.

<div style="border:1px solid">

Every

Can be expressed for all classes by duplication of the adjective construction:
Class. 1: Monna yo mongwe le yo mongwe 'every man'
Class. 3: Monwana wo mongwe le wo mongwe 'every finger'
Class. 5: Lesogana le lengwe le le lengwe 'every young man'
Class 7: Selepe se sengwe le se sengwe 'every axe'
Class 9: Puku ye nngwe le ye nngwe 'every book'

</div>

Figure 7:   Proposed data box for *every* in Sepedi dictionaries

Prinsloo and Gouws (2006) describe this type of repetition of a phrase across the different classes as in figure 7 as grammatical divergence and all such occurences belonging to different classes, e.g. *this man/finger/axe*, etc. or *he/she/ him/her*, etc. could be treated in data boxes with great success.

2.4   Reference to men and women versus addressing them

Groot Noord-Sotho-woordeboek (GNSW) gives the following translation equivalent paradigm for *mohumagadi*: "queen, king's wife, chief's wife, chieftainess, lady, Mrs [a term of courtesy applied to any married woman]", *mohumagatšana*: "miss, queen (of cards)" and for *mosadi*: "woman, married woman, wife". Although all three words refer to a woman/adult female person, the user should be warned that it is inappropriate to address a woman as *mosadi*. The same holds true for a man/adult male person *monna* 'a man' versus *morena* 'Mr.' Consider the suggested data boxes for *women* and *men* in figure 8. This is pragmatic data, a function of data boxes. It is for the lexicographer to decide whether it should be emphasized by inclusion in a data box.

<table>
<tr>
<td>

Although *mosadi, mohumagadi* and *mohumagatšana* all refer to women, do not address a woman as *mosadi* e.g. when greeting her, use *mohumagadi / mohumagatšana*:
Dumela mohumagadi!
Good day madam!

</td>
<td>

Although *monna* and *morena* both refer to men, do not address a man as *monna* e.g. when greeting him, use *Morena*:
Dumela morena!
Good day sir!

</td>
</tr>
</table>

Figure 8:   Proposed data boxes for *addressing females and males* in Sepedi dictionaries

The data box in figure 8 or the applicable sections thereof could be given at the articles of *man* and *woman*.

2.5   Hair

Compilers of Sepedi dictionaries should give clear guidance on the correct meanings and use of Sepedi words dealing with different kinds of hair. ONSD lemmatised *hair* and gives a translation equivalent *moriri*. This is the singular form, i.e. a/one hair. It would be better to give *meriri* 'hair (plural)' as transla-

tion equivalent since in most cases reference to the plural is made. No data box is suggested here, only different treatment of the lemma.

No mention is made of, e.g. the hair of an animal or guidance that *hair* is normally used in the plural form in Sepedi, i.e., *meriri*. GNSW gives the following translation paradigm for *boya*: "hair of an animal, wool, hair of human body (but not of head)" and translates *meriri* as human hair and *mariri* as mane (of a lion). ONSD translates *boya* as wool, animal hair, fur and *mariri* as mane and adding "of a lion" in brackets. *Meriri* is lemmatised but not treated and cross-referred to the singular *moriri* which is simply translated as hair — it should have been translated as human hair. Stronger guidance is required in respect of *hair*, *animal hair*, *mane*, *boya* and *mariri* in order to prevent the user from, e.g. incorrectly using *meriri* to refer to wool or animal hair or *boya* to refer to hair on the head of a person A data box contrasting *boya*, *meriri* and *mariri* as in figure 9 is recommended at the article of *hair*.

Use the correct word in reference to hair
meriri human hair
boya wool, hair of an animal
mariri mane (of a lion)
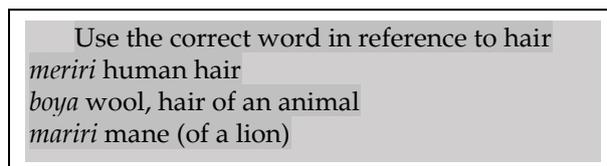
Figure 9:   Proposed data box for *boya*, *meriri* and *mariri*

2.6   Different constructions used for English adjectives

A number of English adjectives such as *last, own, naughty*, etc. are not expressed as adjectives in Sepedi but through different constructions. So, for example, is there no single-word adjective for *naughty* in Sepedi — it is expressed by either a full sentence in the relative mood or by means of a possessive construction as in example (1).

(1)
a. Verbal relative
Mošemane yo a selekago (*mošemane* noun class 1 'a boy' + *yo* demonstrative class 1 + *a* subject concord class 1 + *seleka* verb stem 'be naughty' + *go* relative suffix) 'A naughty boy'

b. Possessive construction
Mošemane wa go seleka (*mošemane* noun class 1 'a boy'+ *wa* possessive concord class 1 + *go* infinitive class prefix class 15 + *seleka* verb stem 'be naughty' 'A naughty boy'

A data box, e.g. as in figure 10 presented at the article of *naughty* will provide the required guidance to the user provided that the target users should have

basic grammatical knowledge of Sepedi. If not, grammatical terms such as *verbal relative* and *possessive construction* should be briefly described in terms of their meaning, i.e. "who is doing something, something of something else respectively". Both options can even be given with the semantic one in brackets, i.e. Verbal relative (who is doing something) and Possessive construction (something of something else).

In Sepedi naughty is not an adjective, there are two ways you can express e.g. **a naughty boy**:

Verbal relative

Mošemane yo a selekago

or

Possessive construction

Mošemane wa go seleka
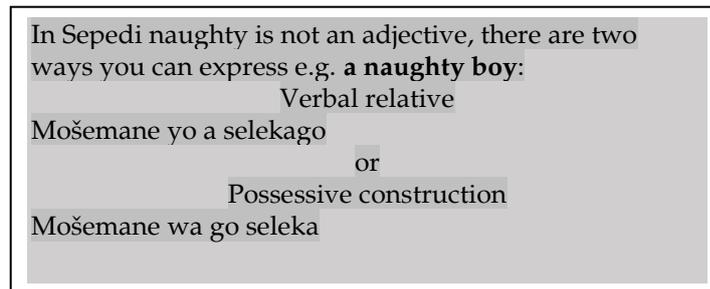
Figure 10: Proposed data box for *naughty*

2.7   Equivalent relations

As a final example consider instances of semantic divergence where a polysemous source language word has more than one translation equivalent (Gouws and Prinsloo 2005). A single Sepedi word *bala* has different translation equivalents, namely *read, count* and *study*. Sepedi has two homonyms *-tala*. The one member of the homonym pair has *old* as its translation equivalent whereas the second homonym has both *green* and *blue* as equivalents. ONSD translates the homonyms *-tala* correctly as respectively *old* and *green, blue*. The user will be well-guided if alerted by means of a data box such as figure 11 because being able to distinguish between *green* and *blue* could be vital in text production situations. A typical situation could be where it is crucial to distinguish between different specific functions performed by e.g. green versus blue buttons on a control panel.

The *Sesotho sa Leboa / English Pukuntšu Dictionary* (SEPD) could mislead the user because only *blue* is given as translation equivalent for *-tala*. The proposed data box in figure 11 could be placed at the article of *-tala* in the Sepedi to English side as well as at the articles for *blue* and *green* in the English to Sepedi side of the dictionary to warn the user that additional clarification might be required in text production situations. It is for the compiler to decide whether sufficient guidance in respect of *-tala* translated as *green* or *blue* was given in the default treatment of the lemma or whether a data box is desired to focus the attention of the user on the different senses. So, for example, the compiler of SEPD will be well-advised to firstly give *green* also as translation equivalent for *-tala*, illustrated by typical examples for each equivalent and further supported by a data box.

> **-tala**: *blue* versus *green*
> Legodimo le le**tala** 'blue sky'
> Bjang bjo bo**tala** 'green grass'

Figure 11: Proposed data box for *tala*

The same holds true for a data box such as figure 12 for *-bala* where the lemma *-bala* has *read*, *count* and *study* as translation equivalents. Although examples will help the user, real success in the treatment of *-tala* and *-bala* would at best be achieved by means of a data box that displays these salient semantic issues.

> **Bala** can mean read
> OR count OR study e.g.
> Ke bala dikuranta 'I read
> the newspapers' or 'I count
> the newspapers' Ba bala
> tikrii ya BA. 'They study for
> a BA degree'

Figure 12: Proposed data box for *bala*

In contrast to one Sepedi word having more than one English equivalent as in figures 11 and 12, a single English word can also have more than one Sepedi equivalent. Sepedi has two words for *ask*, i.e. *botšiša* 'ask, e.g. a question' and *kgopela* 'ask for something'. Consider figure 13 as a data box giving the required guidance on the range of application for *botšiša* and *kgopela*. Such a data box is especially required in the English to Sepedi side of dictionaries such as SEPD where *ask* is simply translated as *botšiša*, *kgopela* without any indication of the range of application.

> In Sepedi there is a difference
> between asking a question and
> asking for something:
>   *Go botšiša potšišo* 'to ask a question'
>   but
>   *Go kgopela tšhelete* 'to ask money'
>   *Go kgopela tshwarelo* 'to ask for
>   forgiveness'

Figure 13: Proposed data box for *ask*

Data boxes are also required at the articles of *wear*, e.g. *apara*/*apere* 'wear clothes' versus *rwala* 'wear a hat' and the many Sepedi equivalents for *close*, e.g. *tswalela* 'close a gate/door' versus *khurumet*ša 'close a container, e.g. the lid of a bottle' versus *khupet*ša 'conceal' versus *moma* 'close(d) mouth', etc.

All these data boxes convey salient information which should be presented to users in a way that draws their attention. The use of data boxes is an ideal presentation method to enable such a transfer of information.

3.   Salient data and more than salient data

Data boxes can be regarded as important and even essential but often under utilized dictionary components which should be used to fulfil a specific need, i.e. typically guiding the user towards carriers of salient data. From the preceding sections it should be clear that data boxes can contribute in a systematic way to assist in the presentation of data that cannot be sufficiently accommodated in the default search positions of dictionary articles or article stretches. The system prevailing in the decision to use data boxes is based on the salience of the specific data. The lexicographic method of using data boxes should not be performed in a haphazard way or as a form of lexicographic face-lifting, cf. Wiegand and Gouws (2011: 238). Lexicographers should have a clear understanding of the reasons why these boxes are used. The presentation of salient lexicographic data can be regarded as one of the major motivations for the use of data boxes in printed dictionaries. Irrespective of what happens in the development of online dictionaries printed dictionaries should preferably continue to use data boxes and even to increase their use. Innovative strategies could complement the traditional way of using these dictionary components. Printed dictionaries of the future could employ data boxes in various ways to respond to new lexicographic challenges.

Data boxes also have an important role to play in online dictionaries. The examples and discussion of data boxes in the following sections of this article should not be regarded as of a language-specific nature but rather relevant to all languages, including the African languages.

Online dictionaries, especially those that were originally planned and published as printed dictionaries, often use data boxes in the same way as found in printed dictionaries. Figure 14 shows the use of a data box as found in the article of the lemma sign *underground* in the OALD to present a usage note: (a comparable usage note is also presented in the articles of the lemmata *metro, subway* and *tube*):

Figure 14: Data box in the article of the lemma *underground* in the OALD

In addition, online dictionaries also display new ways of utilising data boxes. This was already alluded to in Prinsloo and Gouws (2021). This can be seen in figure 15 where the *Merriam-Webster* uses data boxes for navigation in the article of the lemma sign *dull*:

Figure 15: Data box in the article of *dull* for navigation in *Merriam-Webster* for navigation to other search positions

A click on these navigation boxes guides the user to the relevant addresses as seen in figure 16, the address of the link to *Synonyms & Antonyms:*

Figure 16: The address of the link in one of the data boxes in the article of *dull*

The boxes shown in figure 15 do not only have a navigational purpose. They give users the opportunity to unlock textual venues that accommodate additional lexicographic data, as seen in figure 16.

4.   Innovative uses of data boxes in online dictionaries

The transition from printed to online dictionaries can rightly be regarded as extremely important with radical and far-reaching consequences. This can be seen in many aspects of online dictionaries, for example, as indicated by Heuberger (2020: 404), with regard to accessibility of data, multimedia functions, customization, hybridization, user input and storage space. The transition to online lexicography has also had a huge impact on research in the field of metalexicography. Theories of lexicography were primarily developed for the printed environment. The online environment demands a re-assessment of all aspects of these theories, including the various dictionary structures. Some structures of printed dictionaries, for example the article structure, will also prevail in online dictionaries although certain adaptations are needed; some structures, for example the frame structure, are not maintained in online dictionaries. In addition, online dictionaries can also display structures that do not occur in printed dictionaries. An example of such a structure is the screenshot structure, cf. Gouws (2014: 165). When using an online dictionary the user is confronted by various screenshots that are populated by dictionary articles and partial articles. These screenshots display innovative uses of data boxes. With regard to the use of data boxes online dictionaries show that the lexicographic practice has embarked on procedures not yet adequately described or discussed in metalexicographic publications. In the subsequent sections of this article a few occurrences of data boxes in online dictionaries will be identified and briefly discussed in order to show the need for a comprehensive look at data boxes of the future.

4.1   Highlighting data types

Online dictionaries often have dynamic article structures and even multi-layered dynamic article structures (Gouws 2014: 165). The internal access structures provide the user with access routes to the required data in its specific search zone and article layer. This is seen in *elexiko* where the opening screenshot of an article contains data indicators that help the user to move to a next layer of the article and then perhaps to a further layer. When reaching a specific search zone the data indicator as well as the search zone is boxed by means of a thin frame. This frame helps the user to identify the boxed items as a destination of the search route. The following screenshots show this process. Figure 17 is the opening screenshot of the article of the lemma *Arm* (arm) in *elexiko*:
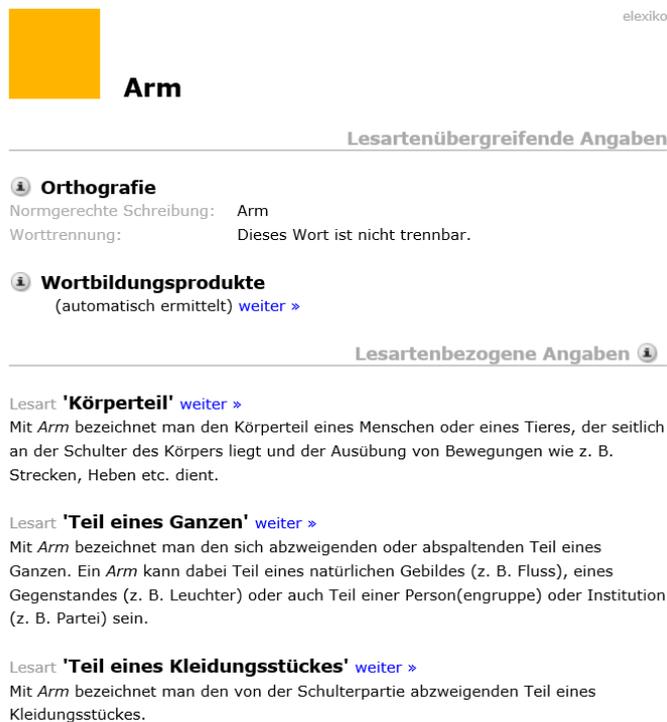
Figure 17: The opening screenshot of the article of the lemma *Arm* in *elexiko*

A user looking for grammatical data regarding the sense of this word referring to a body part finds the data indicator *Körperteil* (Body part) and clicks on the entry *weiter* (=further) next to it. This click moves the user to the partial article presented in figure 18:
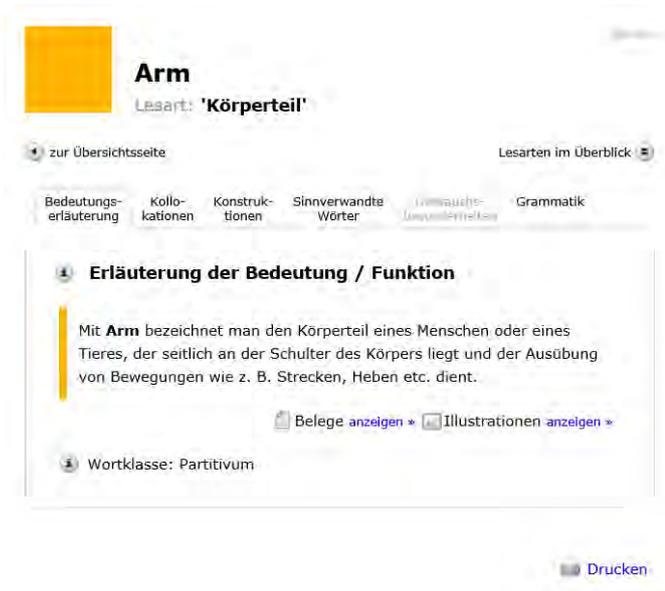
Figure 18: First comment on semantics in the article of the lemma *Arm* in *elexiko*

This screenshot shows the paraphrase of meaning of the specific sense of *Arm* with a thin line putting the paraphrase of meaning, its appropriate data indicator (*Bedeutungserklärung*) (=explanation of meaning) as well as links to example sentences (*Belege*) and illustrations (*Illustrationen*) in a data box. To the right of the data indicator bar the user can find the indicator *Grammatik* (=Grammar) and a click on that indicator opens the next layer, as seen in figure 19:
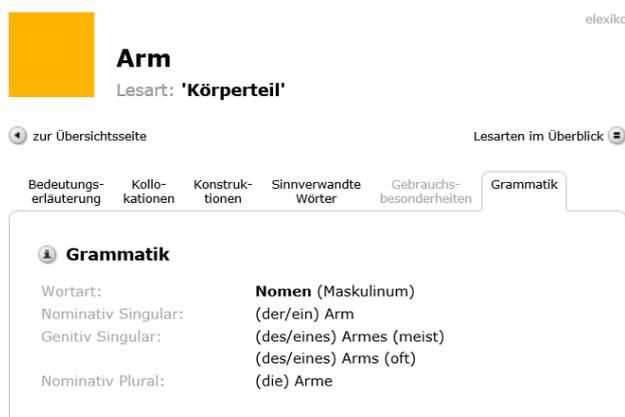
Figure 19: The search zone for grammatical data in the article of *Arm* in *elexiko*

This screenshot shows the grammatical data along with the relevant data indicator appearing in a thinly framed data box.

Both figure 18 and figure 19 display data that are part of the default treatment of nouns in this dictionary. The data box is not used to distinguish salient from less salient data with regard to the presentation in the article as a whole but it does highlight the data salient for the specific consultation — the destination unlocked by the preceding click of a data indicator. This use of data boxes is done in a consistent and systematic way in *elexiko*. It highlights the identification of specific items and enhances the retrieval of the required information. In addition, the type of data box in figure 18 and figure 19 also contributes to improve the layout of the screenshot. This approach is made possible by the dynamic nature of articles in online dictionaries and is in sharp contrast to the limitations due to the static nature of articles in printed dictionaries.

This use of data boxes can also be seen in the following partial article of the lemma sign *koekje* (figure 20) in the Dutch dictionary ANW (*Algemeen Nederlands Woordenboek*). Having navigated from the opening screenshot of the article to the screenshot presenting the partial article in which the sense of *koekje* "small cake" is treated, the user finds a typical partial article layout with three sections presented in columnlike way.

Figure 20: Partial article of the lemma *koekje* in the ANW

The right-hand section contains a data box, unfortunately not as clearly visible in the figure, that accommodates items giving part of speech, spelling and inflection, word relations and pronunciation. In all articles of this dictionary items giving these data types are presented in a data box, situated in the same

position in the screenshot. A similar data box is also seen in figures 21 and 22, giving screenshots with partial articles for two senses of the lemma sign representing the word *muis* (mouse):

Figure 21: Partial article of the lemma *muis* in the ANW

Figure 22: Partial article of the lemma *muis* in the ANW

In figures 20–22 data boxes continue their assignment as containers of lexicographic data but they play an additional role, namely to improve the article

layout and make data easier accessible to the users due to a conspicuous way of presentation. This is a function of data boxes that still needs further exploration. It demands dedicated future work which falls beyond the scope of this article.

4.2   Adding data

4.2.1   Lexicographic data

In online dictionaries data boxes are also used to highlight the access to additional dictionary-internal data that the lexicographer regards as relevant to the word treated in the specific article. Some articles in *Merriam-Webster* have a section "From the editors of *Merriam-Webster*." Below this heading a data box is given in which different types of data can be found. The data are usually of a lexicographic nature and help to fulfil a cognitive function. The data could be a reference to other articles in the dictionary that contain words in the same semantic field as the lemma or it can focus on a discussion of certain related aspect. In figure 23, a screenshot of a partial article of the lemma *bicycle*, shows this data box with its data indicator "10 words every true cyclist will know." A click on this data indicator in the box guides the user to a list of ten articles. This list includes articles with lemmata like *penny-farthing, peloton, velocipede* and *tandem bicycle.*
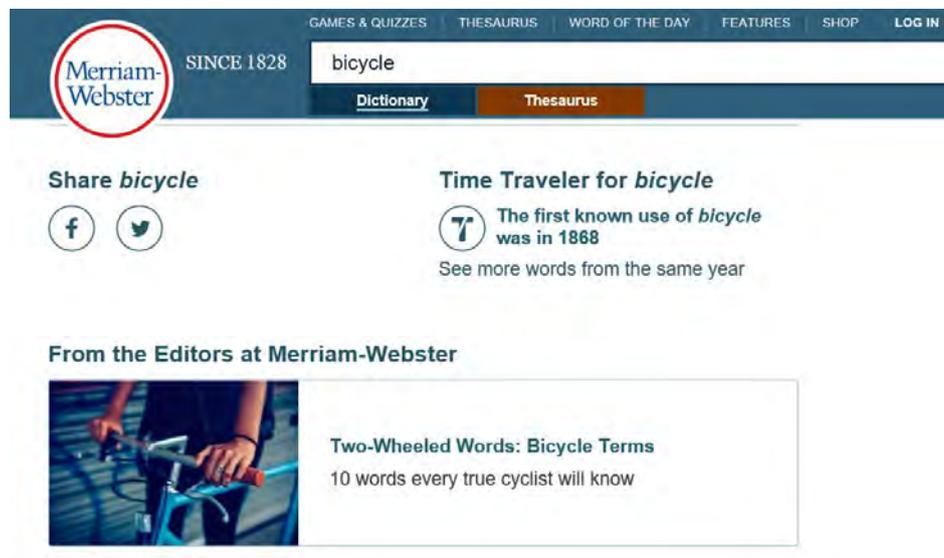
Figure 23: Additional data in *Merriam-Webster*

In this list the article of the lemma *velocipede* has an extensive treatment — with the paraphrase of meaning comparable to that given in the article of the lemma *velocipede*, as seen in figure 24:

Definition of *velocipede*

: a lightweight wheeled vehicle propelled by the rider: such as

a   : TRICYCLE

b   : a 3-wheeled railroad handcar

c   *archaic* : BICYCLE

Figure 24: Paraphrase of meaning of *velocipede* in the MW

The treatment of *velocipede* in the article list is as seen in figure 25:

"Velocipede

Cyclist on Michaux velocipede, 1868

Definition: a lightweight wheeled vehicle propelled by the rider

The predecessor of the pedal bicycle was a two-wheeled vehicle that was propelled with the feet while seated. It was patented in 1818 in Germany, and its name, *Laufmaschine*, ("running machine") was trans-

lated into French as *vélocipède* (from the Latin roots meaning "swift" and "foot"). In English, this contraption was known as the *dandy horse* or the *hobbyhorse*.

Velocipede became a term for any wheeled vehicle propelled by the rider in English, including early bicycles. In French, it became the term used for the improved pedal-powered version designed in the mid-1800s. This was shortened to *vélo* to become their modern word for "bicycle."

The pedal velocipedes were nicknamed *boneshakers* because they were made with wooden wheels and iron frames. After the introduction of rubber tires, boneshaker became a slang term for an uncomfortable or outmoded bicycle.

The *veloc-* of *velocipede* is also the root of *velocity*."

Figure 25: The treatment in the article list of the lemma *velocipede* in the MW

This use of data boxes like that in the article of the lemma *bicycle* shows a significant change in the way in which lexicographers employ this article component — data boxes present a departure slot from where the user can depart to article-external but dictionary-internal data venues. By including these isolated thematically-bound article stretches the lexicographer increases the extent of the dictionary as a search region and the relevant data boxes ensure access to these new venues in the search region.

4.2.2   Non-lexicographic data

Online dictionaries contain typical lexicographic data. Data boxes participate in accommodating the lexicographic data. However, the online environment opens possibilities for dictionaries to become containers of more than just traditional lexicographic data. As components of dictionary articles data boxes in online dictionaries can contain data that even go beyond a display of lexicographic data relevant to the treatment of the word represented by the lemma sign of the specific article. The data distribution structure of these dictionaries can also make provision for the satisfaction of more general cognitive needs. Irrespective of the lemma functioning as guiding element of an article the articles in *lexico.com* contain a data box in which the "word of the day" is given and another box displaying the most recent "word of the year". This is seen in figure 26 with the word of the day in an orange coloured box and the word of the year in a green coloured box:
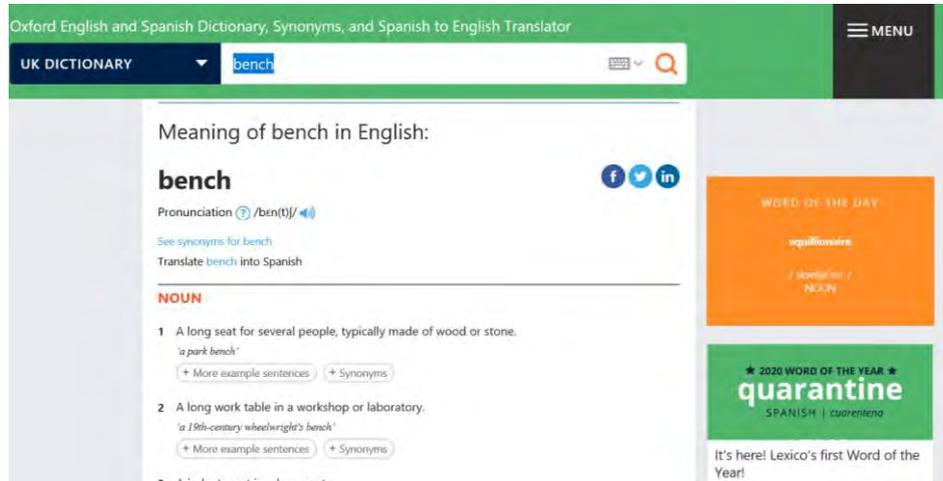
Figure 26: Word of the day in *lexico.com*

Because these boxes are presented in every article, knowledgeable users of this dictionary will know that they can retrieve this information from the dictionary and know where to find it. For a user consulting the dictionary for the first time or consulting it to find other data in a dictionary article these data boxes offer a data bonus and additional consultation success.

Lexicographers of online dictionaries also use the lesser space restrictions to include data boxes with non-lexicographic data that could be seen as a type of lexicotainment, where *lexicotainment* could refer to the presentation of data that do not contribute to achieving the genuine purpose of the dictionary, but enable the retrieval of information that might not be lexicographically relevant but may enrich the consultation procedure. Schierholz (2015: 340) also refers to "reading dictionaries for entertainment or to kill time (which is called 'lexico-tainment'". The following screenshot of a partial article of the lemma *bench* in *lexico.com* shows a data box that contains a brief quiz of which the topic is not related to the lemma of the article accommodating this data box. On any given day this quiz will not be the same in all articles. However, the subsequent data box with "trending words" is the same in all articles. The data in this latter data box are not actually a form of lexicotainment because this box given in figure 27 rather adds to the fulfilment of a cognitive function of the dictionary and therefore this data fall within the scope of the genuine purpose of the diction-ary.
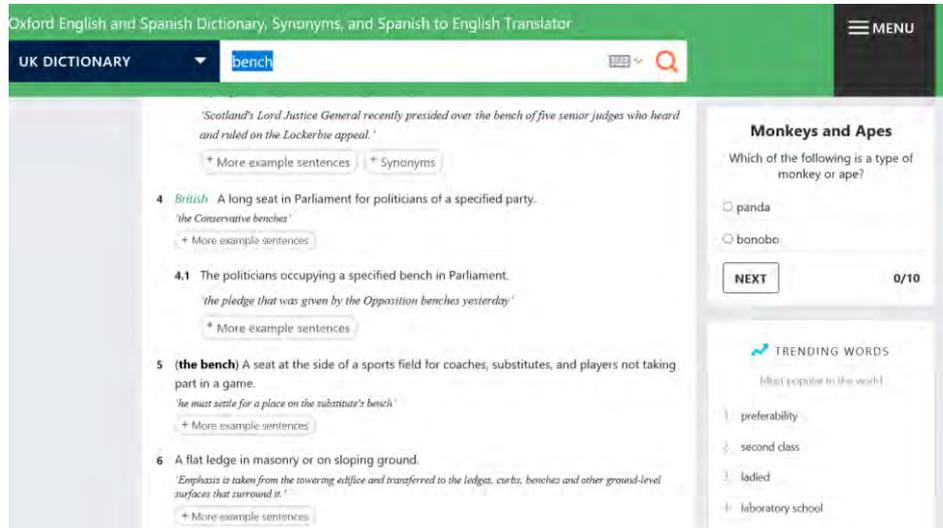
Figure 27: A quiz and "Trending words" in *lexico.com*

Lexicographers can also employ data boxes to respond to questions from their dictionary users. Articles in the *Merriam-Webster* dictionary contain a data box "Ask the editors". This data box given in figure 28 contains separate boxes with the response of the lexicographer to questions put by the users:
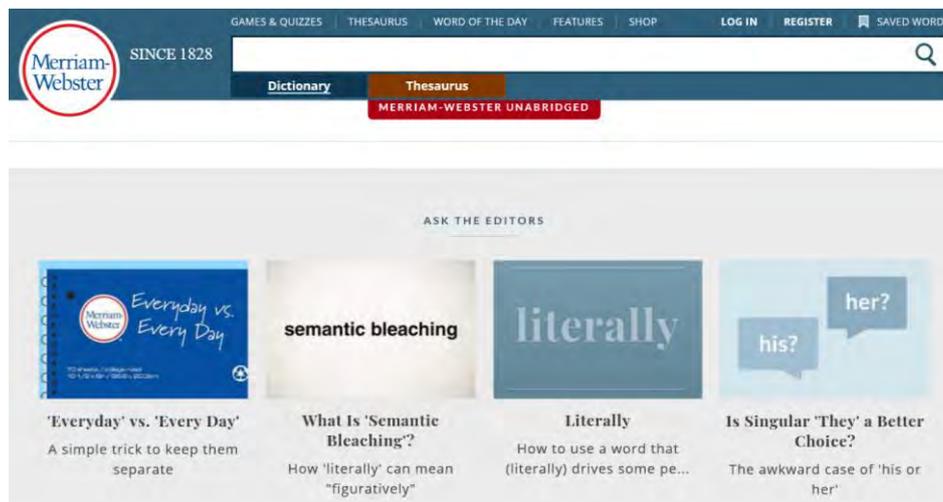
Figure 28: Response of editors to user questions

This data box is used to introduce an innovative communication opportunity between dictionary maker and dictionary user. This use of data boxes and the further possibilities that could arise demand more comprehensive attention from the field of metalexicography.

4.3   Information and reference tools

As utility tools dictionaries, whether in printed or online format, are carriers of data from which users can retrieve information. Online dictionaries are no longer only regarded as isolated tools but they are part of a larger family of reference tools. Besides presenting lexicographic data to their users online dictionaries often also guide the users to dictionary-external sources — either in the same search domain, a dictionary portal, or in the search universe where other lexicographic and non-lexicographic sources can be targeted. Although the mediostructure of printed dictionaries also makes provision for cross-reference positions accommodated by cross-reference items with a dictionary-external address these cross-references typically are embedded within the dictionary article — either within a search zone complementing another item or in a search zone dedicated to dictionary-external cross-references, as seen in figure 29, the article of the lemma *Benutzungsgrund* in the *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research* (WLWF: Wiegand et al. 2010)
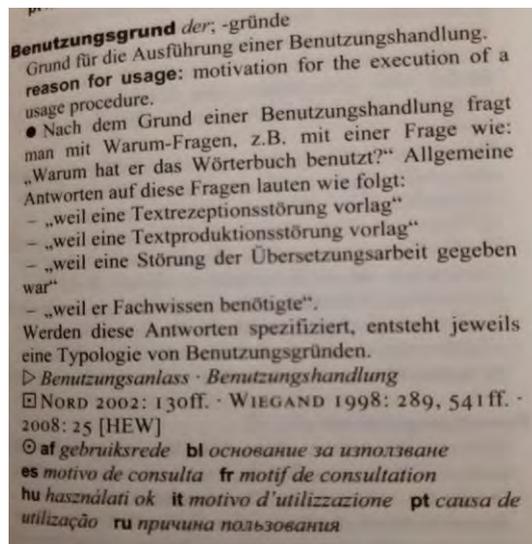
Figure 29: Article of the lemma *Benutzungsgrund* in the WLWF

In this article the typographical structural indicator "■" identifies the search zone populated by items giving dictionary-external cross-reference addresses, as seen in figure 30:
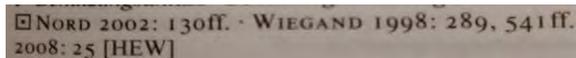
Figure 30: Dictionary-external cross-reference addresses in the WLWF

Online dictionaries can cross-refer users in a much better way to a specific reference address, for example by including a link with an unambiguous data marker, as seen in figure 15. It can also be done by directing users to sources in either the same search domain or in the search universe (Gouws 2021: 15; 2021a). Unlike presenting these sources as items in a search zone populating the obligatory microstructure of the dictionary, as seen in figure 29, the lexicographer can use a data box that contains, among others, a reference to different sources from which the user can retrieve additional information. This is seen in figure 31, the article of the lemma sign *Zug* in *dict.cc* where a click on the information icon in the left and the right margins of the article activates a pop-up data box, seen in figure 31 in the bottom right-hand corner of the article.
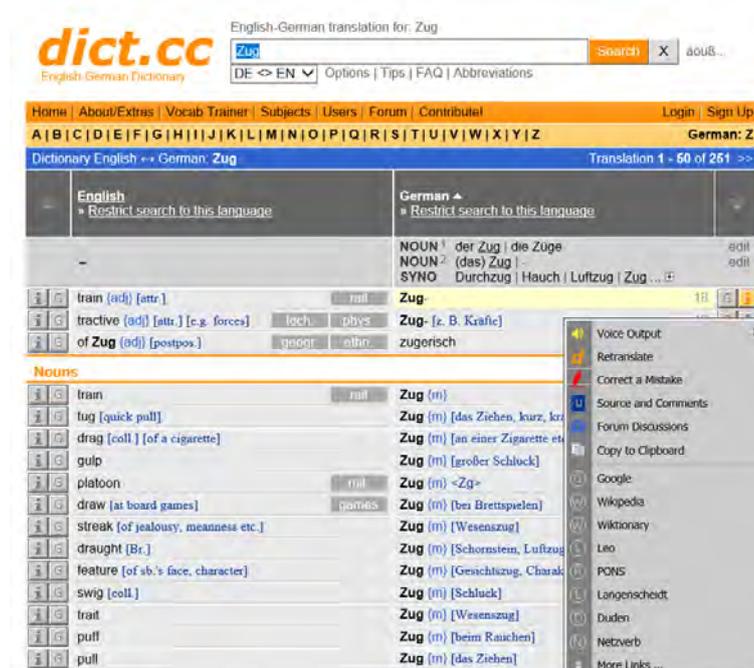
Figure 31: Article of the lemma *Zug* in *dict.cc*

The lower section of this data box is used to convey another type of salient data, namely the titles of dictionary-external sources. A click on any of these sources guides the user to the treatment of the item of which the information icon was clicked in the source given in the data box. Here the data box assists the user in a way not typically found in printed dictionaries.

This is another innovative use of data boxes. The dictionary introduces an occurrence of this type of text constituent that gives access to data relevant to the lemma and it also gives access to other sources where additional relevant information could be retrieved.

5.   The future

Data boxes have made a significant contribution in ensuring a more comprehensive and diverse transfer of lexicographic data. Lexicographers of both printed and online dictionaries have been innovative in introducing different ways of best employing data boxes. Certain procedures, for example the procedure of boxing salient data, became established in printed dictionaries. This tradition has been continued in some online dictionaries. The reality of lesser space restrictions but also dynamic article structures, new layout possibilities and easier linking of items in a dictionary article to either dictionary-internal or dictionary-external addresses have resulted in new ways of using data boxes in online dictionaries. Many of these ways have not been sufficiently discussed in metalexicographic literature and this paper emphasises the need for such a discussion. Not only lexicographic data but also relevant non-lexicographic data can be accommodated in data boxes. This offers numerous opportunities to lexicographers when devising the data distribution structures of their dictionaries. Much more attention can now be given to the possibility of a stronger focus on the cognitive function of dictionaries.

Electronic dictionaries of the future are expected to continue the tradition of the paper and current electronic dictionary to present data as part of the treatment of the lemma. So, for example in figure 14 the data box is presented directly following the treatment of the second sense in the article of *underground*. In this way the users have no option whether they want to see the data box or not. Presentation of databoxes in this way can add to information overload and increase text density. Hyperlinking could be a better or alternative approach to the presentation of data boxes in future electronic dictionaries. Electronic dictionaries employ hyperlinking and pop-up boxes to such an extent that almost every item in a dictionary article is hyperlinked to a pop-up box. Such pop-up boxes provide the user with information on various issues ranging from convention explanation, phonetic and grammatical information and translation equivalents or individual words used in paraphrase of meaning; thus, a complicated cross-referencing system. This system is designed on the basis of two approaches namely hovering and clicking. In the case of hovering no deliberate action from the user is required but an opportunity is

offered to them to obtain more information through a deliberate clicking action. Consider an inventory of pop-up boxes obtained through hovering and clicking for *mosadi* compiled by Prinsloo and Van Graan (2021: 54) in figure 32.
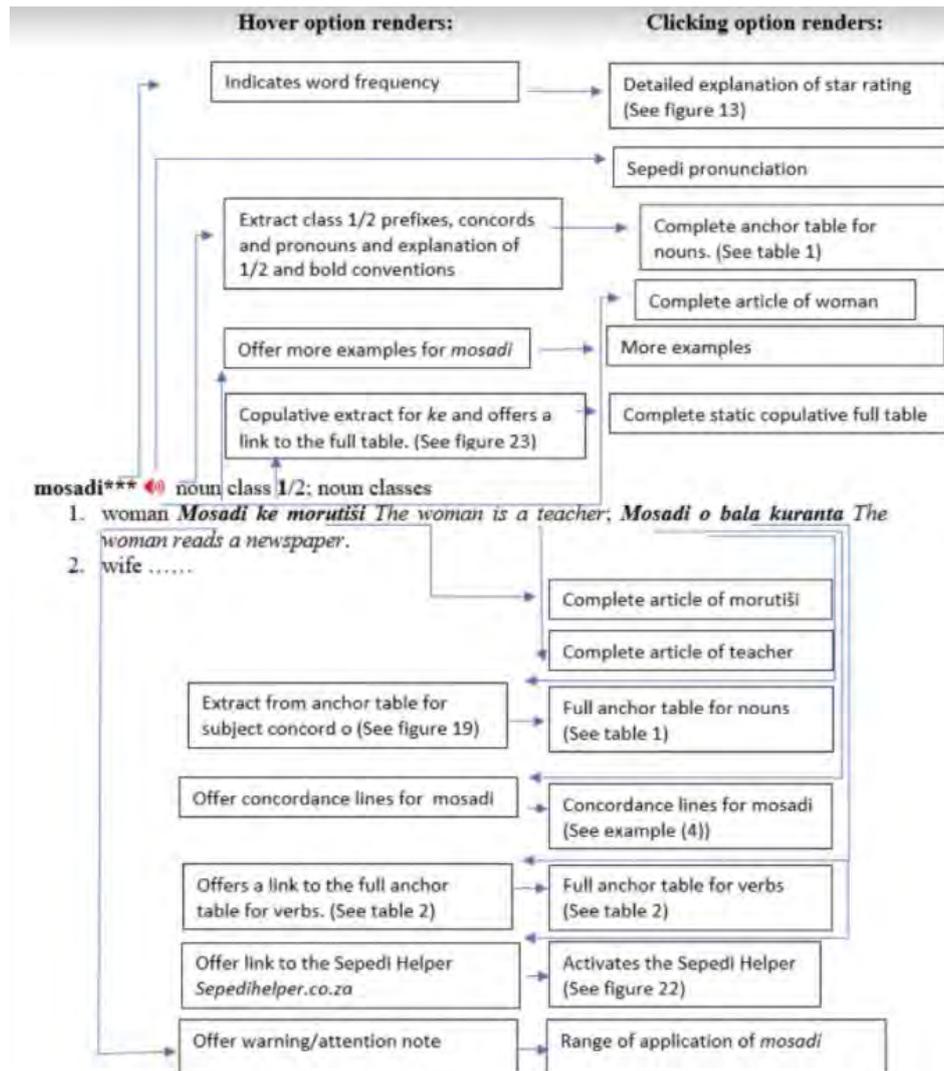
Figure 32: Model article for *mosadi* with arrows indicating hovering and clicking options

The first 20 pop-up boxes deal with a variety of issues such as frequency indication, pronunciation, grammatical guidance, additional examples, complete

articles of words used in the translation equivalent paradigm, etc. The final two pop-up boxes are data boxes offering salient information pertaining to the range of application of *mosadi*. A hierarchy exists between these two data boxes. Through hovering over the warning/attention note as is the case in the top left box on frequency, the user is informed about a box that can be obtained through clicking. Two considerations are at stake here. Firstly, the issue of information overload and secondly, a hierarchical drill-down strategy. Prinsloo and Bothma (2020: 87) say in this regard:

> A user does not need such an information overload to solve a very specific information need in a given situation – the user typically prefers to be provided with exactly the required amount of information to solve his/her information need in the given situation. In an e-environment, this information overload can easily be circumvented by initially providing only basic information that builds upon existing knowledge, but then providing, through drill-down options on demand, either more basic or more in-depth information about the problem at hand.

Drill-down actions through clicking in figure 32 lead to another (deeper) level of information. In the case of frequency indication the drill-down action renders a pop-up box with detailed information on the star rated convention used for frequency indication in the dictionary. In the *Macmillan Dictionary* (MED) such drill-down actions result in the provision of a wealth of information for the user. Hovering over the frequency star convention in any dictionary article guides the way to several levels of drill-down options. The first level is detailed information "RED WORDS AND STARS". Second levels obtained through further clicking are clicking on a video entitled "Smart learning with Red Words and Stars" and a clicking option to download a "Red Words & Stars pack". The same holds true for the data box "Range of application of *mosadi*" in figure 32 where the drilling-down action results in data boxes such as the one designed for *mosadi* in figure 8 above.

One of the exciting possibilities in online lexicography is the use of data pulling procedures (Gouws 2018; 2021). The successful employment of data pulling procedures can be enhanced by a clear indication of the information retrieval structure of the specific dictionary. In this regard it is important that users need be made aware of the dictionary-external sources functioning in the relevant search domain as well as the search universe. Data boxes can make a huge contribution in presenting a position in a dictionary article where the menu of dictionary-external sources can be given — as seen in figure 31.

In further metalexicographic research into data distribution options in dictionaries as well as into the enhanced use of data pulling procedures an increased use of data boxes should be negotiated. This is a dictionary component that could continue to play a significant role in future dictionaries.

6.   Conclusion

This article as well as the preceding two articles in this trio have focused on a variety of aspects related to data boxes in printed and online dictionaries. The first article (Gouws and Prinsloo 2021) gave a metalexicographic perspective with a focus primarily on the occurrence of lexicographic data boxes as text constituents in dictionaries. In the second article (Prinsloo and Gouws 2021) the types and contents of data boxes were discussed. This third contribution put the emphasis on data boxes in bilingual dictionaries with an African language as one of the treated languages. It also looked at new ways in which existing online dictionaries have used data boxes.

The current use of data boxes in both printed and online dictionaries can form an important point of departure for the future use of this type of text constituents. **Accommodating salient data should remain a significant assignment to data boxes**. In addition, the use of data boxes to ensure an improved article layout and data distribution gives future lexicographers numerous options to enhance the quality of their dictionaries. As dynamic utility tools dictionaries can also use data boxes as text constituents that form a bridge between dictionary-internal and dictionary-external consultation procedures.

Data boxes have played a significant role in the lexicographic practice. This role should be maintained and increased in future dictionaries. Better collaboration between metalexicographers and practical lexicographers can ensure an exciting use of data boxes when fully exploiting the potential of the online environment.

Endnote

1.   The term 'Bantu' got stigmatized during the Apartheid Era in South Africa. Therefore, the term 'African' is preferred in South Africa even in reference to what is internationally referred to as 'Bantu languages'. The discussion in this article is, however, focused on the Bantu language family and most of the issues described cannot necessarily be generalized to be applicable to other languages on the continent of Africa. To respect the view of those opposed to the term 'Bantu', it will only be used in cases where a distinction between African languages (languages spoken in Africa) versus a member of the Bantu language family is essential.

Acknowledgement

References

Dictionaries

ANW. *Algemeen Nederlands Woordenboek.* Available at http://anw.inl.nl/. (Date of access: June 2021.)

dict.cc. Online-Wörterbuch Englisch–Deutsch/Deutsch–Englisch. Available at https://www.dict.cc/. (Date of access: June 2021.)

elexiko. Online-Wörterbuch zur deutschen Gegenwartssprache. Available at https://www.owid.de/docs/elex/start.jsp. (Date of access: June 2021.)

GNSW = Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho-woordeboek / Comprehensive Northern Sotho Dictionary / Pukuntšu ya Sesotho sa Leboa.* Pretoria: J.L. van Schaik.

lexico.com. Oxford English and Spanish Dictionary, Thesaurus, and Spanish to English Translator. https://www.lexico.com/.

MED = *Macmillan Dictionary.* Available at https://www.macmillandictionary.com/. (Date of access: August 2021.)

Merriam-Webster. Available at https://www.merriam-webster.com/dictionary/. (Date of access: June 2021.)

OALD = *Oxford Advanced Learner's Dictionary.* Available at http://oald8.oxfordlearnersdictionaries.com/dictionary/underground_3. (Date of access: June 2021.)

ONSD = De Schryver, G.-M. (Ed.). 2007 *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

OZSD = De Schryver, G.-M. (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford.* Cape Town: Oxford University Press Southern Africa.

SEPD = Mojela, M.V., M.C. Mphahlele, M.P. Mogodi and M.R. Selokela. 2006. *Sesotho sa Leboa / English Pukuntšu Dictionary.* Cape Town: Phumelela.

Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer, and W. Wolski (Eds.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research.* Volume 1. Berlin: De Gruyter.

Other references

Gallardo, A. 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography:* 59-69. Columbias: Hornbeam Press.

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

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

Gouws, R.H. 2021. Expanding the Use of Corpora in the Lexicographic Process of Online Dictionaries. Piosik, M., J. Taborek und M. Woźnicka (Eds.). *10. Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven:* 1-19. Berlin/Boston: De Gruyter.

Gouws, R.H. 2021a. Lexicography and Documentation in a Multilingual Environment. Unpublished paper delivered at *the 14th International Conference of the Asian Association for Lexicography, ASIALEX 2021, 12–14 June 2021.*

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: SUN PReSS.

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

Heuberger, R. 2020. Monolingual Online Dictionaries for Learners of English and the Opportunities of the Electronic Medium: A Critical Survey. *International Journal of Lexicography* 33(4): 404-416.

Lombard, D.P. 1985. *Introduction to the Grammar of Northern Sotho.* Pretoria: J.L. van Schaik.

Louwrens, L.J. 1991. *Aspects of Northern Sotho Grammar.* Pretoria: Via Afrika.

Poulos, G. and L.J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho.* Pretoria: Via Afrika.

Prinsloo, D.J. 2017. Africa's Response to the Corpus Revolution. Xu, Hai (Ed.). 2017. *Proceedings of the 11th International Conference of the Asian Association for Lexicography, ASIALEX 2017, 10–12 June 2017, Guangzhou, China: Lexicography in Asia: Challenges, Innovations and Prospects:* 20-31. Guangzhou, China: ASIALEX. Available at http://asialex.org/pdf/Asialex-Proceedings-2017.pdf. [Consulted January 2021.]

Prinsloo, D.J. 2020. Detection and Lexicographic Treatment of Salient Features in e-Dictionaries for African Languages. *International Journal of Lexicography* 33(3): 269-287.

Prinsloo, D.J. and T.J.D. Bothma. 2020. A Copulative Decision Tree as a Writing Tool for Sepedi. *South African Journal of African Languages* 40(1): 85-97.

Prinsloo, D.J. and R.H. Gouws. 2006. Lexicographic Presentation of Grammatical Divergence in Sesotho sa Leboa. *South African Journal of African Languages.* 26(4) 184-197.

Prinsloo, D.J. and R.H. Gouws. 2021. Lexicographic Data Boxes. Part 2: Types and Contents of Data Boxes with Particular Focus on Dictionaries for English and African Languages. *Lexikos* 31: 374-401.

Prinsloo, D.J. and N.D. van Graan. 2021. Principles and Practice of Cross-referencing in Paper and Electronic Dictionaries with Specific Reference to African Languages. *Lexicography, Journal of Asialex* 8(1): 32-58.

Prinsloo, D.J. and J.J. van Wyk. 1992. Verwantskapsterminologie van die Noord-Sotho. *South African Journal of Ethnology* 15(2): 43-58.

Schierholz, Stefan J. 2015. Methods in Lexicography and Dictionary Research. *Lexikos* 25: 323-352.

Van Wyk, E.B., P.S. Groenewald, D.J. Prinsloo, J.H.M. Kock and E. Taljard. 1992. *Northern Sotho for First-Years.* Pretoria: J.L. van Schaik.

Van Wyk, J.J. and F.T. Haasbroek. 1990. Verwantskapsterminologie van die Batswana. *South African Journal of Ethnology* 13(4): 159-179.

Wiegand, H.E. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Volume 1.* Berlin/New York: Walter de Gruyter.

Wiegand, H.E. and R.H. Gouws. 2011. Theoriebedingte Wörterbuchformprobleme und wörterbuchformbedingte Benutzerprobleme I. Ein Beitrag zur Wörterbuchkritik und zur Erweiterung der Theorie der Wörterbuchform. *Lexikos* 21: 232-297.

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Sestiende Deel: SRP–SZONDITOETS.* 2021, xx + 759 pp. ISBN-13 978-1-990998-46-1 (leerband), ISBN-13 978-1-990998-45-4 (plastiekband). Stellenbosch: Buro van die WAT. Prys: R900 (leerband) / R550 (plastiekband).

Inleidend

By die terhandneming van WAT XVI (2021), is 'n mens bewus van die historiese waarde van hierdie laaste gedrukte oplaag, 'n indrukwekkende foliant wat 'n voortsetting is van die formaat wat mettertyd gevestig geraak het as kenmerkend van hierdie vlagskippublikasie. Hierdie derde en laaste deel van die letter S strek van die lemma **SRP**, die afkorting vir *sekuriteitsreguleringspaneel*, tot die lemma **Szonditoets**. Aangesien die oorblywende dele ná S slegs aanlyn beskikbaar sal wees, word die papier-era hiermee afgesluit en die elektroniese medium ten volle ingespan. Deur voortgesette hersiening en bywerking, wat hierdeur moontlik gemaak word, bly die WAT 'n aktuele bron van kennis oor die Afrikaanse woordeskat. Hierdie uitgawe beslaan 779 (759 + xx) bladsye, met 7 606 verklaarde lemmas in die sentrale teks. Hierby kom ook 'n addendum wat ongeveer 1 560 bygewerkte lemmas van **AAHS** tot **Sri Lankaans** bevat, en dus prakties deel vorm van die sentrale teks.

Vorderingstempo

Die eerste 15 dele (met 11 bykomende oplae) van die WAT het teen 'n frekwensie van gemiddeld 4 jaar en 10 maande per deel sedert 1950 van die pers gekom.

Dl I (1950)	Dl IX (1994)
Dl II (1955)	Dl X (1996)
Dl III (1957)	Dl XI (2000)
Dl IV (1961)	Dl XII (2005)
Dl V (1968)	Dl XIII (2009)
Dl VI (1976)	Dl XIV (2013)
Dl VII (1984)	Dl XV (2019)
Dl VIII (1991)	

Deel sewentien (T) word vroeg in 2023 in die vooruitsig gestel, en die mikpunt vir die voltooiing van die woordeboek tot Z (ten minste voor 'n beoogde volledige herbewerking) is nou 2028. Dit is moontlik gemaak deur bykomende befondsing en personeeluitbreiding, waardeur die beplande voltooiingsdatum met 10 jaar vervroeg is. Vanaf 2020 het die effek van die heersende pandemie terselfdertyd maatreëls genoodsaak om die voorgenome voortgang van die werk te verseker, ten spyte van afsonderlike werkstasies.

Beoordelingsperspektief: Die WAT as omvattende woordeboek

'n Belangrike faktor by die beoordeling van enige woordeboek is die aard en doel van die betrokke publikasie, wat soos volg lui:.

> Die missie van die WAT is om 'n omvattend verklarende woordeboek en ander leksikografiese produkte in Afrikaans saam te stel, en sodoende die Afrikaanse taal, sy gebruikers en die leksikografie te dien.

As teikengebruikers word uit die breë spektrum dan veral "hoërskoolleerders tot Afrikaans-akademici" uitgesonder, vir wie die aanbiedingswyse toeganklik behoort te wees. 'n Verdere eienskap wat beklemtoon word, is dat dit 'n sinchroniese woordeboek is, en dus nie aanwysings van die herkomstyd of geskiedenis bevat nie, en voorbeeldmateriaal ook nie chronologies gerangskik is nie. Dit beteken egter nie dat ouer (en verouderende) vorme nie opgeneem word nie — wel dat sulke vorme deur etikette gemerk word.

In die onderstaande bespreking word daar eers 'n kort oorsig gegee van die hoofkomponente, waarna bepaalde kenmerke van die hoofteks bespreek en ook met mekaar in verband gebring word.

Die voorwerk

In die voorwerk word omvattende inligting verstrek, in aansluiting by die reeds bekende formaat van WAT XV en die voorafgaande dele. Daarby word 'n beskrywing van die mediostruktuur, of die wyse waarop kruisverwysings en bronverwysings aangebied word, aangevul deur 'n wye verskeidenheid aktuele inligting, wat die produksieproses, ander leksikografiese produkte, die personeel, skakeling met deskundiges en die publiek, en die missiestelling van die WAT (hierbo vermeld) insluit. 'n Enkele punt van kritiek wat betref die indrukwekkende lys vakkundige medewerkers is dat postume bydraers nie as sodanig aangedui word nie — o.a. vdr. B. Hinwood en mnr. C. de Ruyter, wat beide al 'n geruime tyd oorlede is. Of dit wel ter sake is, is wel 'n ander vraag, maar dit sou waarskynlik nodig wees om nuwe deskundiges te betrek vir die toekomstige dele.

'n Eienskap van die toeligting (in die voorwerk) by die gebruik van die woordeboek is die uitvoerige beskrywing van die leksikografiese bewerking — iets wat vir die gebruiker grafies voorgestel word aan die hand van die (nou reeds bekende) skematiese voorstelling van inskrywings (hier op bl. vi).

Die makrostruktuur

Die feit dat die WAT 'n omvattende woordeboek is, beteken dat die makrostruktuur (d.w.s. die geheel van leksikale items wat as lemmas ingesluit word) 'n aansienlike mate van vryheid kan weerspieël wat betref die keuse van lem-

mas, en by alle waters kan skep.

In 'n onderhoud op RSG op 6 Junie 2021 wys Alet Cloete daarop dat die omvattende aard van die makrostruktuur weerspieël word deur die opname van historiese of verouderde woorde, nuutskeppings, vaktaal, formele, informele en plat vorme, en alle variëteite, wat streektaal- en geselstaalvorme insluit. Die insluiting van al hierdie kategorieë, asook betekenisonderskeidings wat met verloop van tyd verandering ondergaan, veroorsaak dat daar tydens die samestelling van die betrokke deel ordeningsbesluite geneem moet word op grond van die huidige gebruiklikheid. (Hierop word daar verder in die bespreking uitgebrei.)

Die mikrostruktuur

Wat die mikrostruktuur van artikelinskrywings betref — dus die wyse waarop die inligting oor elke lemma gestruktureer is — word aspekte soos grammatiese inligting, lettergreepverdeling, spelling, uitspraakleiding, woordsoortelikheid, morfologie, etimologie en sintaksis gedek. Hierdie struktuur maak voorsiening vir 'n breë spektrum van moontlikhede, vanaf 'n enkele item (bv. 'n afkortingsvorm soos "**Swaz.** afk. Swaziland" tot 'n omvang van 5 200 woorde (in die geval van **steek**), of selfs langer. Oor die rangskikking van die elemente wat deel kan vorm van die mikrostruktuur word in die toeligting altesaam 12 kategorieë bespreek, wat kumulatief bydra tot die mees volledige struktuur van 'n artikel, maar wat uiteraard selde almal aan die beurt kom. Ook aspekte soos die aanduiding van samestellings en afleidings, verwysings en wisselvorme kom ter sprake.

Bylaes

Benewens die lys van 55 bladsye met bronne waaruit in die artikelinskrywings aangehaal is, voorafgegaan deur 'n toeligtingsbladsy, is daar 'n enkele bylae (Addendum A, wat ongeveer 1 560 nuwe byvoegings vanaf Deel I tot XV insluit).

Die deskriptiewe en preskriptiewe aard van die woordeboek

As naslaanbron het die WAT nie 'n eksplisiet preskriptiewe funksie nie, maar weerspieël dit die breeds moontlike spektrum van geskrewe Afrikaans. Hoewel dus oorwegend deskriptief, sou die aanbieding van alle leksikografiese teks (bv. lemmas, definisies en mikrostrukturele inligting) wel ortografiese AWS-norme volg wat op Afrikaans van toepassing is. Vergelyk egter, by wyse van uitsondering, die lemmas **sulu** en **Sulu**, waarvan die meervoudsvorme as **sulus** en **Sulus** (sonder meervoudsapostroof) aangedui word.

Ten spyte van die feit dat die WAT teoreties die grense oorskry van wat as ortografies genormeerde Afrikaans beskou kan word (wat betref die weergawe van sitate), word daar in die omgang deur die deursnee-taalgebruiker dikwels na hierdie bron verwys om 'n bepaalde gebruiksvorm as gemagtig of gesanksioneer te staaf. Definisies van lemmas in die WAT as verklarende woordeboek word uiteraard wel as gesaghebbend aanvaar en stawend gebruik, 'n funksie wat daartoe lei dat die WAT deur die gewone gebruiker ook oor 'n breër spektrum van normering gebruik word.

Lemmatisering van fonetiese spelwyses (wat uitspraakverskynsels weerspieël)

Deur die opname van sitate uit variëteite wat (nog) nie ortografies genormeer is nie, bv. uit Kaaps, kom daar variante spellings voor wat dikwels op die fonetiese interpretasie van individuele skrywers berus. Dit is in sulke gevalle waar die sitaat, en in sommige gevalle die lemma wat daardeur geïllustreer word, 'n tentatiewe normerende funksie het. 'n Voorbeeld van 'n geselstaalvariant, wat ook in verskillende variëteite voorkom, is **spieg** (p. 701), wat eerder 'n ontronde geselstaalvariant is van **spuug**, waarna daar nie verwys word nie, as van die meer neutrale **spoeg**. 'n Vergelykbare voorbeeld van die lemmatisering van fonetiese spelwyses is die opname van **spien** (met 'n kruisverwysing na [2]SPEEN), 'n enkele voorbeeld uit 'n veel groter aanbod in die sitate, soos *ernstagge, 'ie, vloe, gie, suste, dieselle wôt, vrint, vanaan, nuh, ôs*, ens., wat almal fonologiese kenmerke van Kaaps weerspieël). Dit onderstreep die behoefte aan ortografiese normering ten opsigte van veral informele variëteite. Daar sou aan die hand gedoen kon word dat so 'n tentatiewe normering kan berus op 'n frekwensiegebaseerde aanduiding in die etikette, waardeur daar met groter stelligheid 'n keuse uitgeoefen kan word van hoe relatiewe frekwensie beskryf word. Dit het bv. betrekking op die onderskeid tussen etikette soos "meer dikw." (by **stêre** en ook **stêrre**), "ongewoon" (by **sterde**), en "selde ook" (by **sterde** en **stêre**), waarna hieronder verwys word.

Etikette — motivering en toepassing

Deur die gebruik van etikette word sommige geselekteerde lemmas gemerk wat benewens 'n basiese (denotatiewe) betekenis, ook spesifieke konnotasies dra waarvan die gebruiker kennis behoort te dra. Sulke etikette kan moontlik in algemene kategorieë verdeel word:

(a) Register, wat deur die stylvlak bepaal word, of die mate van formaliteit — redaksioneel word hierna as sosiostilistiese etikette verwys (Voorbeelde: *geselstaal, skertsend, formeel, verhewe*, ens.);
(b) Streeks- of groepsaanduidend (Voorbeelde: *streektaal, visserstaal, studentetaal, sleng*);
(c) Tegniese of vakterminologie (Voorbeelde: *anatomie, oudheidkunde, skeepvaart, stylleer, kookkuns, chemie*);

(d) Stigmatisering (Voorbeelde: *vloek*, *plat*, *vulgêr*, *neerhalend*, *rassisties*, *seksisties*, *skeltaal* — as teenhanger van sulke terme word die etiket *eufemisties* ook gebruik);

(e) Gebruiklikheid en tydsgebondenheid (Voorbeelde: *minder gebruiklik*, *ongewoon*, *selde ook*, *meer dikwels*, *verouderd*, *verouderend*, *histories*);

(f) Inligting wat nie kontekstueel uit die sitaat of poëem afgelei kan word nie, soos by **suursmaak**, hoewel die sitaat hier deur die gebruik van "(*sauerkraut*)" reeds voldoende inligting bevat.

Die gebruik van etikette om registers, variëteite e.d.m. aan te dui, is 'n noodsaaklike komponent van die mikrostruktuur, gegewe die seleksie uit alle soorte variëteite waaruit sitate onttrek word, en kan van groot waarde wees om die variasiemoontlikhede van die taal leksikografies te struktureer. Aan die ander kant het sulke etikette ook die funksie om die beperkings ten opsigte van register en konteks onder die aandag te bring. By die bewerking van die betrokke lemma word die kontekstuele beperkings ook, al is dit implisiet, deur die sitate geïllustreer.

Een kategorie uit die lys etikettipes hierbo word met meer as gewone aandag onder die loep geneem, naamlik stigmatiseringsetikette. Omdat taal soveel aspekte van die gemeenskaplike bestaan van 'n taalgemeenskap (en verskillende taalgemeenskappe in 'n land soos Suid-Afrika) benoem, sluit dit ook die problematiek van sodanige saambestaan in. In taalkommunikasie dien die woord dikwels ook as wapen, en kom veral as uitdrukking van rassistiese en seksistiese houdings in die taalgebruik van sommige gebruikers na vore. Terwyl die bestudering van sulke items wetenskaplik onproblematies kan wees, kan die opname van emosioneel gelaaide items vir die deursnee-gebruiker dui op 'n erkenning van die gebruik daarvan, en dus as genormeerd (soos hierbo vermeld). 'n Duidelike uiteensetting van die opname van sulke items (benewens die gebruik van etikette) word in par. 6 van die voorwerk verstrek, iets waarvan die gebruiker met vrug kan kennis neem.

Opmerkings oor ortografiese aanbiedingswyse van lemmas

Wat betref die skryfwyse van samevoegings wat los of vas geskryf kan word, veral dan as bywoord plus voltooide deelwoord, sou die een en ander oor die aanbiedingswyse opgemerk kan word. Ook hier geld die waarneming dat vorme wat in die WAT gelemmatiseer word, deur heelwat sprekers as norm aanvaar word. In die geval van voorbeelde soos

sterkontwikkel, sterkruikend, sterkgewortel, sterkgroeiend

word die vas geskrewe (in die teks *vasgeskrewe*) vorme wel gelys op grond van die voorkoms daarvan in sitate, maar die etiket lui "(*meer dikw. los*)", wat by korpusondersoek dikwels blyk slegs los te wees. In die definisie van **sterkontwikkel** lui dit ook (los) "Wat sterk … ontwikkel is". Die indruk kan hierdeur

geskep word dat slegs die attributiewe vorme vas geskryf word, bv. ook **sterk-skemer**, en die predikatief los. Die antonimiese vorme met **swak** (bv. **swakont-wikkelde**, maar ook **swak ontwikkelde**, wat beide wel in korpora voorkom) volg dieselfde patroon, waar **swak-** as koppeltekenlemma aan die begin van 'n lys onverklaarde lemmas geplaas word, met 'n etiket wat aandui dat dit meer dikwels los geskryf word. Daardeur kan die afleiding (tereg of ten onregte) ook gemaak word dat daar in 'n mate selektief omgegaan is met die keuse van lemmas wat deur konvensie of semantiese verbleking (naas die los geskrewe by-woord plus naamwoord, wat uiteraard nie gelemmatiseer hoef te word nie) as vaste vorme kan voorkom.

In teenstelling hiermee word die mees gebruikte vorm, meestal los geskrewe, by idiomatiese en gespesialiseerde uitdrukkings (wat ook vas geskrewe teen-hangers met 'n laer frekwensie het) as basis gebruik (vgl. die tweede alinea van par. 5.3 op p. xvii).

Slot

WAT XVI voeg in verskeie opsigte waarde toe aan sy voorgangers, o.a. deur die uitgebreide terminologieverklaring, die insluiting van klassieke (bv. Romaanse of Grieksgebaseerde) ekwivalente van Germaanse vakwoordeskat, en duidelike uitspraakleiding. Met inagneming van bepaalde kritiese opmerkings, kan hier-die sestiende deel beskou word as 'n uitstekende bron vir taalpraktisyns, navorsers en studente. Die redelik gekompliseerde mikrostruktuur maak dit ook van nut vir meer gevorderde gebruikers, en voldoen aan leksikografiese vereis-tes op 'n hoë vlak. Dit kan beskou word as 'n model vir kennis oor omvattende, verklarende woordeboeke.

Die grootste uitdaging wat betref die afhandeling van die woordeboek as geheel is waarskynlik om 'n konsekwente formaat van aanbieding van A tot Z in die uiteindelike aanlyn weergawe tot stand te bring.

Ernst Kotzé
Emeritus
Nelson Mandela-Universiteit
Port Elizabeth
(ernst.somerstrand@gmail.com)

Another isiXhosa Monolingual Dictionary: A Critical Analysis of the *Shuters IsiChazi-magama SesiXhosa*[*]

Dion Nkomo, *School of Languages and Literatures:
African Language Studies, Rhodes University, Makhanda (Grahamstown),
South Africa (d.nkomo@ru.ac.za)*

Abstract: This article offers a critical analysis and evaluation of the *Shuters IsiChazi-magama SesiXhosa*, the second monolingual dictionary in the history of isiXhosa. The analysis draws theoretical and methodological insights from dictionary criticism as a domain of metalexicography and analyses the dictionary in view of existing dictionaries in the language, especially the first monolingual dictionary published under the auspices of the IsiXhosa National Lexicography Unit (XNLU) a decade earlier. Concise paraphrases of meaning are identified as the major strength of this dictionary when compared to its XNLU predecessor. However, inconsistent provision of microstructural entries, an imbalanced macrostructure and a non-integrated front matter reduce the user-friendliness of the dictionary. Nevertheless, another dictionary in isiXhosa produced without the involvement of a National Lexicography Unit is welcome towards the intellectualisation of African languages.

Keywords: AFRICAN LANGUAGES, DICTIONARY, DICTIONARY CRITICISM, INTELLECTUALISATION OF AFRICAN LANGUAGES, *ISICHAZI-MAGAMA*, ISIXHOSA LEXICOGRAPHY, MACROSTRUCTURE, MICROSTRUCTURE, NATIONAL LEXICOGRAPHY UNITS, FRONT MATTER TEXTS

Isishwankathelo: Esinye isiChazimagama sesiXhosa esiLwiminye: UHlalutyo lwesiChazimagama sakwaShuter and Shooter. Eli nqaku lihlalutya i*Shuters IsiChazi-magama SesiXhosa*, esisesesibini isichazimagama ukuba lwiminye kwimbali yesiXhosa. Olu hlalutyo lulandela ithiyori neenkqubo zokuhla amahlongwane izichazimagama njengommandla wofundonzulu ngezichazimagama. Uhlalutyo luhlahlela esi sichazimagama ngokusithelekisa nezichazimagama ezikhoyo esiXhoseni ingakumbi¸ *IsiChazi-magama SesiXhosa* esaqulunqwa liZiko lesiZwe loChazo-magama lwesiXhosa kwiminyaka elishumi ngaphambili, esisesokuqala ukuba lwiminye esiXhoseni. Iinkcazelo zeentsingiselo ezithe ngqo zichongwe njengezona zinempumelelo kwesi sichazimagama xa sithelekiswa nesasandulelayo. Nangona kunjalo iinkcukacha ezinikezelwa ngemichazwa azinikwa ngokuchanekileyo kananjalo uluhlu lwemichazwa alusibonisi kakuhle isakhiwo sesigama sesiXhosa.

Oku nokungayondelelani kweziqulatho zesiqu sesichazimagama kunokwenza kungabi lula ukusisebenzisa isichazimagama. Nangona kunjalo ukupapashwa kwesinye isichazimagama esiXhoseni okungabandakanyi iZiko lesiZwe loChazomagama kwamkelekile kuphuhliso lweelwimi zesiNtu.

Amagama angundoqo: IILWIMI ZESINTU, ISICHAZIMAGAMA, UHLALUTYO LWEZI-CHAZIMAGAMA, UPHUHLISO LWEELWIMI ZESINTU, *ISICHAZI-MAGAMA SESIXHOSA*, UCHAZOMAGAMA LWESIXHOSA, ISAKHIWO SOLUHLU LWEMICHAZWA, ISAKHIWO SEENKCUKACHA NGEMICHAZWA, AMAZIKO ESIZWE OCHAZOMAGAMA, UMPHAMBILI WESICHAZIMAGAMA

1. Introduction

In 2018, a decade after the publication of the first monolingual isiXhosa dictionary, namely *IsiChazi-magama SesiXhosa*, edited by S.L. Tshabe, Z. Guzana and A.B.B. Nokele under the auspices of the IsiXhosa National Lexicography Unit (XNLU) in 2008, Shuter and Shooter published another monolingual isiXhosa dictionary edited by M.M.M. Duka. Only a fine print *Shuters* before the title of the new dictionary distinguishes the names of the two monolingual isiXhosa dictionaries. Duka also claims that his dictionary, just like its predecessor, is for use at all educational levels, including higher education, as well as by the generality of isiXhosa speakers. Specific reference is made to the study of isiXhosa as a subject and its use as a medium of teaching. The publication of a second monolingual dictionary in isiXhosa could therefore be seen as another milestone in the intellectualisation of the language. Kaschula and Nkomo (2019) identify lexicography as a vitally important corpus planning enterprise in the intellectualisation of any language. There is a definite relationship between language policy and lexicography (Gouws 2007a; Nkomo 2018). The relationship is also similar between language policy and other language intellectualisation activities such as the development of orthographies, terminology development, translation and language teaching/learning (Kaschula and Nkomo 2019).

Gouws (2007a) illustrates the influence of language policy on lexicography in South Africa from a historical perspective. The legacy of apartheid history is evident when one considers the lexicographic infrastructure, practices, products and dictionary culture in African languages compared to Afrikaans and English. Some African languages largely remained in the pre-lexicographic era (Gouws and Ponelis 1992; Nkomo 2020) well into the democratic era, while others only had missionary efforts to thank for the few dictionaries that existed (Gouws 2007a; Nkomo 2020). Despite the development of orthographies and terminologies, as well as the production of textbooks for basic education and creative literary works during the missionary and apartheid periods, limited functional spaces crippled the standardisation of indigenous African languages and, hence, their lexicography. Accordingly, the state of lexicographic practice in the languages validates the remarks made by Gallardo (1980) when he writes:

> … only fully standardized languages have their lexicon organized in monolingual dictionaries. … In non-standardized language situations, dictionaries do not exist or, at best, are bilingual dictionaries, that is, compiled in function of a different language. … apart from being always bilingual, [the dictionaries] are usually not compiled by members of the speech community involved, who are not even able to use them (Gallardo 1980: 61).

This is true for isiXhosa in which dictionaries such as Kropf's *A Kafir–English Dictionary* of 1899, McLaren's *A Concise Kafir–English Dictionary* (1915) and the *Oxford English–Xhosa Dictionary* (Fischer, Weiss, Mdala and Tshabe 1985) were produced, but never a monolingual dictionary. The compilers of those dictionaries were also consistently clear about their target users, i.e. non-mother-tongue learners of isiXhosa who comprised early missionaries deployed in the Eastern Cape and later on English-speaking learners of isiXhosa in formal education. Such a situation prevailed in African languages lexicography in general as reported in Awak (1990) and Busane (1990), among others.

The elevation of the nine previously marginalised indigenous African languages into official languages alongside Afrikaans and English culminated in, among other developments, the establishment of the National Lexicography Units (NLUs). Alberts (2011) outlines the main function of the NLUs as the compilation of comprehensive general-purpose monolingual dictionaries. This focus recognises the role of monolingual dictionaries in the intellectualisation of languages. Such a role has been recognised in the transformation of European vernaculars that were inferior to Latin and Greek into powerful modern languages such as English, French and Italian (Nkomo 2018). Compared to the situation described by Gallardo (1980), as cited above, monolingual dictionaries are a case of *lexicography for the language speakers by the language speakers themselves*. At a time when the NLUs are experiencing challenges that threaten their mandate and long-term future, dictionaries such as the *Shuters IsiChazi-magama SesiXhosa*, henceforth the Shuters dictionary, produced outside the NLUs as the official lexicographic entities are vitally important. However, also of equal importance is their quality and overall lexicographic contributions in the development of the respective languages. The present article seeks to contribute in this respect by describing, analysing and evaluating the Shuters dictionary within the realm of dictionary criticism, the theoretical and methodological perspectives of which are outlined after a brief overview of the dictionary in the next section.

2.   An overview of *the Shuters IsiChazi-magama SesiXhosa*

In a celebratory preface, the lexicographer exalts the value of his dictionary, particularly that it is monolingual when he writes:

> Lo mqulu, ISICHAZI-MAGAMA SESIXHOSA, usisiphumo soluvo lokuba *isiXhosa lulwimi oluvuthiweyo olukwisigaba sokusetyenziswa kuqulunqwe isiChazi-magama sesiXhosa poqo.* Kungoko iintsingiselo zamagama, kulo mqulu, zicaciswe zachazwa ngesiXhosa kuphela (own emphasis in italics).
>
> This compilation, ISICHAZI-MAGAMA SESIXHOSA, is a result of the view that *isiXhosa is a fully-fledged language that is at the level of being used in the compilation of an exclusively isiXhosa dictionary.* Accordingly, the meaning of words is this dictionary are described and explained only in isiXhosa (Duka 2018: v).

Duka is a language enthusiast, an author and an experienced educator who describes himself as an ardent dictionary user (Duka 2021).[1] As a student of languages and literature, he studied English, Latin and Afrikaans in the 1960s before going on to get a doctorate degree in African Languages. He recalls with fondness the pivotal role played by dictionaries in his education, and attributes some teaching/learning challenges and a poor dictionary culture among speakers of South African languages to the paucity of dictionaries of diverse types (cf. Nkomo 2020). Accordingly, he conceived his dictionary with a view to contribute not only towards the learning and study of isiXhosa among its mother-tongue speakers, but also in consolidating its use in the teaching of other academic subjects, particularly Mathematics and Science, of which he is a retired teacher and chief education specialist.

In terms of size, the Shuters dictionary is a small dictionary with 237 pages of the main text and 16 pages of front matter texts (see Section 5.3 below). The dictionary is just less than half the size of its XNLU predecessor, whose main text is 520 pages. In terms of its contents and envisaged role, the blurb text states that:

— Sicacisa amagama ngendlela elula etsho kuthi dlwe.
  (It explains words in a simple way that clarifies meaning).

— Sinika iseseko esiqinileyo nesomeleleyo ekubhaleni nasekuthetheni isiXhosa esifanelekileyo.
  (It provides a solid foundation for writing and speaking 'appropriate' isiXhosa).

— Sidwelisa amahlelo ezibizo ngendlela elula neyamkelekileyo.
  (It outlines noun classes in a simple and accepted manner).

— Sakha ubungqondi kumfundi.
  (It builds understanding in the learner/student).

— Sichaphazela izifundo ngezifundo.
  (It covers different academic subjects).

— Silungele ukusetyenziswa ezikolweni nakumabanga aphezulu.
  (It is appropriate for use in schools and higher levels of education).

Such a text informs potential buyers and users about what they should expect from the dictionary. However, such information is conceived from the compiler's perspective (Hartmann and James 1998). Swanepoel (2017b) notes that blurb texts are prepared by dictionary publishers quite often to promote sales. The user perspective and critical perspectives are necessary to offer a balanced assessment and estimation of the value of dictionaries. Dictionary criticism, discussed in Section 3 and Section 5, respectively provides the relevant theoretical and methodological guidance for this undertaking.

3.   Dictionary criticism: A metalexicographical perspective

Hartmann and James (1998: 32) state that dictionary criticism is "concerned with the description and evaluation of DICTIONARIES and other reference works". A mutual relationship exists between dictionary criticism and meta-lexicography or theoretical lexicography. Wiegand (1984) allocated space for dictionary criticism in his general theory of lexicography at the nascent stages of theorisation of lexicography as a discipline. Writing from the perspective of the function theory, Tarp (2017: 116) posits that "[c]riticism of existing dictionaries should be regarded as an important area of lexicographical research". Its rationale taps deeply from the fundamental importance of dictionaries as functional tools and lexicography as a problem-solving practice. This means that dictionary critics must have noble motivations. Gouws (2017: 40) makes this point when he writes:

> Dictionary criticism is not done for the mere sake of criticizing but in order to inform and to improve — to inform dictionary users and to help lexicographers and publishers to improve the quality of future dictionaries.

Nielsen (2009: 25) and Gouws (2017) extend Wiegand's (1984) notion of the *genuine purpose* of dictionaries to dictionary criticism. If dictionary criticism is to achieve its genuine purpose, then it should be "free from bias" (Nielsen 2009: 37) arising from "ethically dubious" motivations noted by Tarp (2017: 117), such as "to humiliate its author, to promote one's own dictionary, to lavish praise on a friend, or to please an authority in order to get personal benefits". Such temptations may arise from a narrow interpretation of the practice, with a potential of equating dictionary criticism to dictionary bashing. Accordingly, some scholars such as Nielsen (2009) and Swanepoel (2017a; 2017b) prefer *dictionary reviewing* and *dictionary reviews* as alternative terms to refer to the practice and its products respectively, both of which ought to yield a "true and fair view of the dictionary concerned " (Nielsen 2009: 36).

   In order for dictionary criticism to be fit for purpose, dictionary critics need not only to have good intentions as those posited by Gouws (2017). The dictionary critics also need to be fit for purpose. They must possess "knowledge of lexicographic theories, principles and practices" (Nielsen 2009: 29). In an unpublished paper, Hadebe (2005: 2) bemoaned the quality of some dictionary

reviews because they approached dictionaries "as if they were school textbooks or grammar books and at worst as if they are works of fiction". This happens too often in the case of reviews written for newspapers by journalists with no basic training in lexicography, be it practical or theoretical.

Lexicographic practice in African languages currently appears to benefit less from sound theoretical guidance in the form of lexicographic research and dictionary criticism. Practicing lexicographers, especially from the NLUs in South Africa, are under extreme pressure to meet unrealistic demands by the Pan South African Language Board (PanSALB) that they should publish at least one dictionary every year in spite of inadequate financial, human and technological resources. They therefore struggle to interact not only with the academic research in order to learn from or influence it, but also with the language-speaking communities who are the potential users of their products. As such, dictionaries are produced in African languages with potential users taking very little, if any, notice, while lexicographers get little, if any, useful feedback in the form of dictionary criticism. To paraphrase Samuel Johnson's words of disillusionment, lexicographers become truly harmless drudges whose preoccupation and outputs bear little significance in their communities (cf. Mugglestone 2015: 1).

The poverty of dictionary criticism manifested itself in the aftermath of the publication of the *Oxford Bilingual School Dictionary: IsiXhosa and English* in 2014. The dictionary generated passionate discussions in newspapers, radio and social media platforms (Nkomo 2015). The discussions were particularly characterised by sensational and scathing remarks from disappointed mother-tongue speakers of isiXhosa because of missing words, the inclusion of loanwords, typically from English, at the expense of indigenous coinages, e.g. *igranti* instead of *indodla* (social grant), and some missing senses. Particularly incredible were recommendations for the isiXhosa-speaking community to boycott the dictionary. Remarkably, such protests were made at the neglect of the overall potential of the dictionary to address the lexicographic needs experienced by learners of both isiXhosa and English in the language and content subjects. Since discussions involved largely language speakers, including journalists and the so-called social influencers, without sound lexicographic grounding, they were classified as "common conversations" (Nkomo 2015: 43). As researchers and scholarly products, dictionary critics and their works are expected to be "in a different league than ordinary people and ordinary products" (Nielsen 2009: 25). Nkomo (2015: 45) attributes uninformed and unconstructive dictionary criticism to a lack of a developed societal dictionary culture. In their *Dictionary of Lexicography*, Hartmann and James (1998: 41) define *dictionary culture* as "[t]he critical awareness of the value and limitations of dictionaries and other reference works in a particular community". In the context of a poor or non-existent societal dictionary culture, community members may not only be unaware of existing dictionaries, but they may also be unaware of the value of dictionaries and the qualities of a good dictionary. Thus, Gouws (2017: 33) argues that "dictionary criticism should be part of a comprehensive dictionary

culture". By subjecting the Shuters dictionary to critical evaluation, this article therefore hopes to contribute towards efforts of establishing a dictionary culture among isiXhosa speakers and speakers of other African languages which are being further intellectualised through lexicography, in the same way as the 'Towards One-Learner One-Dictionary' Project in selected Eastern Cape schools (Nkomo 2020).

4.   A methodological approach to dictionary criticism

Swanepoel (2017b: 13) makes an important point that dictionary reviewing is "a research process". Every research process needs to be carefully designed with a clear set of methodical activities in order to generate relevant data sets and analyse them systematically using appropriate theories. Dictionary criticism as research is not exempted from this requirement. Thus, Gouws (2017: 34) writes:

> Although dictionaries are and should be the target of numerous and diverse reviews, the criticism should not be done in a haphazard way. A scientific approach, even for a review targeted at a small audience is needed.

Svensén (2009) identifies two general methods that are used in dictionary criticism, namely the desktop method and the test method. Their main distinction is that the former evaluates the dictionary by simply analysing the dictionary while the latter evaluates the dictionary based on experimental dictionary usage by selected targeted users. Since dictionaries are utility products compiled to meet specific needs of specific users in practical situations, the latter is undeniably superior to the former. Nevertheless, the desktop method is also capable of generating sound and evaluative conclusions, as long as it is undertaken systematically and consistently with theoretical and practical purposes of dictionary criticism. At that level, desktop dictionary evaluation is akin to literary criticism that is conducted independent of text reception tests being done with real readers of literary works. What is important is that the criticism and evaluation offer useful insights for the producers and consumers of texts. However, for dictionaries the emphasis needs to be placed on their functionality as utility products.

Nielsen (2009) makes a distinction between maximising and minimising dictionary evaluation. The former "attempts to give as exhaustive description, analysis and evaluation of a dictionary as possible" while the latter "is deliberately limited to selected aspects regarding the dictionary" (Nielsen 2009: 32). This article attempts a maximising approach as it reflects on the macrostructure, the microstructure and outer texts in relation to the supposed functions of the dictionary. This approach is capable of yielding a fair reflection of the dictionary, which is possible when the critic follows the guidance of Gouws (2017) in conducting:

> … assessment of … [the] dictionary that reflects both positive and negative features with regard to the macro- and microstructural coverage and its presentation, the data distribution, the dictionary structures, the satisfaction of the envisaged lexicographic functions, the response to the intended target users' needs and reference skills, and to make recommendations for the improvement of the specific dictionary, and ways to make an optimal retrieval of information from the data on offer (Gouws 2017: 38).

However, one needs to heed an important guiding principle for desktop dictionary criticism as inspired by Svensén (2009), i.e. that "a reviewer cannot read, analyze, describe and evaluate any dictionary from beginning to end but necessarily has to make an informed choice out of all the design features under review" (Swanepoel 2017a: 23). Guided by this principle, the present paper adopted the following procedures out of Svensén's (2009: 484-485) comprehensive list:

— Dictionary familiarisation which included reading the blurb and all the introductory texts of the dictionary in order to determine the dictionary functions

— Browsing the main text of the dictionary to determine how it seeks to achieve its functions

— Scanning through the macrostructure of the dictionary to determine the length of alphabetic stretches and the methods of ordering

— Random selection of lemmata to determine their types and lemma selection criteria

— A close study of the articles of lemmata falling under different speech categories to determine the data items and data indicators that constitute the typical microstructural design of the dictionary

It is important to note that these activities were not conducted in a purely chronological order. For example, while studying the dictionary macrostructure, front matter texts referring to specific types of lemmata or grammatical categories in isiXhosa were consulted to corroborate some observations regarding relevant statements of intent made in the front matter texts. This was helpful in distinguishing systemic issues from possible editorial mistakes in the dictionary. Although the intention was never to read the entire dictionary as advised by Svensén, the iterative process was sufficient to get a fair picture of the dictionary in view of its brief overview provided above, while leading to evaluative comments made in the remainder of the article. At the same time, the Shuters dictionary was also compared with other isiXhosa dictionaries in order to determine how it addressed certain issues differently from, or similarly to, its predecessors.

5.   A critical analysis and evaluation of the *Shuters IsiChazi-magama SesiXhosa*

Building on the compiler's perspective in Section 2 and the two sections that followed, this section attempts a critical analysis and evaluation of the Shuters dictionary. This undertaking is important in the light of the responsibility that lexicographers have towards society. As Gouws (2017: 26) writes, "[l]exicographers and their works should be judged to determine if they have successfully complied with this assignment" of serving society. Dictionary critics, thus, complement practicing lexicographers who produce dictionaries for specific communities or users within the communities, by making the communities aware of the dictionaries, their strengths and their shortcomings. To do that, it becomes imperative for the dictionary critic to delve deeper into the dictionary beyond the introductory and blurb texts. As per advice from Nielsen (2009: 25), "dictionary reviews should not merely describe but should contain more or less thorough analyses, evaluations and reflections". Accordingly, the three subsections of this section analyse and evaluate the Shuters dictionary in terms of its macrostructure, microstructure and outer texts.

5.1   The macrostructure

On the level of macrostructure, at least three major questions characterise dictionary criticism. The first question pertains to the size of the lemma stock of the dictionary. The second question focuses on the components and composition of the lemma stock. The final question considers the arrangement of lemmata and its user-friendliness. By addressing such questions, a dictionary can be evaluated in terms of whether it can address the needs of users who may consult the dictionary with respect to their vocabulary needs and any other relevant data that constitutes lexicographic treatment. This section of the paper seeks to elaborate on the first two questions and address them by giving examples from the Shuters dictionary in order to evaluate the extent to which the dictionary can or does assist its (identified) target users. The arrangement of lemmata will be addressed anecdotally in Section 5.2.

5.1.1   Size and composition of macrostructure

As noted in Section 2, the Shuters dictionary is a modest dictionary in terms of its size. According to Duka (2021), the publisher stipulated a limit of 3000 lemmata for this dictionary. Based on this figure, one may wonder about the adequacy of the macrostructure in addressing vocabulary needs of quite a broad spectrum of target users, namely basic and higher education, not to mention the generality of isiXhosa speakers. No indication is particularly made regarding the priority of any of those user categories.

In order to determine the macrostructural adequacy of the Shuters dictionary, the composition of the macrostructure was considered. When confronted with a rigid lemma or page limit for a dictionary, the decision of what to include or exclude ought to have been more critical for the lexicographer. Lemma selection needed a clear set of criteria informed by the user profile and the needs of users. This would ensure that every lemma would be worth its space in the dictionary.

Furthermore, the criteria adopted needed to be applied against a dictionary basis from which lemma candidates are identified. Traditionally, existing dictionaries and other published materials such as literary works, textbooks and print media have constituted dictionary bases for new dictionaries. With advances in computational linguistics nowadays, such texts can be assembled into electronic corpora from which lemmata may be drawn using corpus query software such as Wordsmith tools or Sketch Engine. Unfortunately, building adequate African languages corpora for lexicographic purposes is still confronted with challenges in terms of size, balance, representativeness and copyright issues. This undermines the general advantages of using corpora for as dictionary bases.

For the Shuters dictionary, the lexicographer acknowledges a few texts as constituting his dictionary basis. These include *IsiXhosa seBanga leMatriki* (Pahl et al. 1967), *Nasi isiXhosa: Ibanga* 7 (Sigcu and Manyase 1981), The *Grammar of isiXhosa* (Oosthuysen 2016), *Introduction to the Phonology of the Bantu Languages* (Meinhof and van Warmelo 1932) and Bongela (1991)'s *Amagontsi*, which are grammar books, as well as an isiXhosa poetry anthology entitled *Isihobe se-Afrika Entsha* (Duka, Mhlontlo and Matoto 2010). Added to these texts are three dictionaries, namely *A Concise Xhosa–English Dictionary* (McLaren 1915), the *Oxford English–Xhosa Dictionary* (1985) and the *Zulu–English Dictionary* (Doke and Vilakazi 1948). The lexicographer does not indicate his motivation for choosing the above texts for his dictionary basis. What is notable is that, apart from the texts by Meinhof and van Warmelo (1932) and Oosthuysen (2016), most of the rest of the grammar books are for basic education. This suggests the prioritisation of the school market ahead of university students and the generality of isiXhosa speakers. What is also remarkable is the exclusion of the XNLU *IsiChazi-magama SesiXhosa*, hitherto the only monolingual dictionary compiled for mother-tongue speakers, in favour of those compiled for non-mother tongue speakers of the language decades ago. Overall, the use of rather old texts creates a potential chasm between the dictionary and its target users, particularly the school market where new contemporary texts are used. Given the small size of the Shuters dictionary, the XNLU dictionary is most likely to remain the more prestigious, authoritative and contemporary of the two monolingual dictionaries. However, as noted earlier, this is in the context of a poor dictionary culture.

Without any cue in the preface or introduction of the dictionary, further analysis of lemma selection and composition of the macrostructure relies on

insights gained from the interview conducted with the lexicographer. Duka (2021) confirmed an intuitive approach to lemma selection, arguing that he knew what he needed to include in the dictionary as a mother-tongue speaker, a writer and an education specialist. A close study of lemmata throughout the dictionary was also conducted to complement the discussion with the lexicographer. Given the modest size of the dictionary, this was not a toll order, and it provided an overview of the composition of the macrostructural entries of the Shuters dictionary.

One important point to note regarding the macrostructure of the Shuters dictionary is a conspicuous word-bias. Semi-lexical items and multi-lexical items were not entered in the dictionary. In an agglutinating language such as isiXhosa, semi-lexical items include prefixes, suffixes and other formatives that make the language morphologically highly productive. Treating such formatives as lemmata would work in a complementary way in serving the cognitive and text production functions with the grammatical notes, which describe the structure, and functions of isiXhosa nouns, verbs, pronouns, adjectives, copulatives, etc., between pages vi-xvi. Multi-lexical items would include proverbs and idioms, which are only included as microstructural entries under specific lemmata. For example, the proverb 'Ucuntsu akafani noshici' is included under the lemma -**cuntsu**, but not as a lemma. This strong word-bias is, however, understandable against the strict prescription of a 3000 word-limit macrostructure.

Existing literature has noted a bias towards nouns and verbs as lemma candidates in Bantu language dictionaries (De Schryver 2008; De Schryver and Wilkes 2008). The bias is ill informed as it is only justified by the dominance and majority of nouns and verbs as content words in the vocabulary of all languages. De Schryver and Wilkes (2008: 828) astutely remind practicing lexicographers that:

> One cannot use a language without the content words from other word classes, however, and certainly not without function words which glue all content words together.

Accordingly, "an approach which throws out most word categories" (De Schryver and Wilkes 2008: 831) reduces the utility value and user-friendliness of dictionaries. Instead, De Schryver (2008: 66) argues for "a clear approach to the treatment of each and every … word class". Such an approach does not prevail in the Shuters dictionary, in which lemma selection was guided by the lexicographer's intuition. Nouns and verbs dominate the dictionary at the expense of other word categories. To illustrate this bias, the two longest alphabetic stretches are letters *Nn* and *Bb*, which have a combined total of six hundred and eight lemmata. Four hundred and sixty-four (76.3%) of the lemmata are nouns, while one hundred and sixty-nine (27.8%) are verbs and only forty-four (7.2%) are drawn from other categories. The implications may not be as catastrophic as lamented by De Schryver and Wilkes (2008), given that this dictionary is for

mother-tongue speakers, but the inclusion of adjectives and quantitative pronouns would benefit some dictionary users. This could be done without consuming much dictionary space, as De Schryver (2008: 67) estimates that Bantu languages have about twenty to thirty so-called 'true adjective stems'". Most examples that are listed in two front matter texts entitled 'Izimelabizo' (Pronouns) and 'Izichazi' (Adjectives) are not even lemmatised in the dictionary.

The lexicographer's emphasis of the value of his dictionary as an educational tool, especially to support the use of isiXhosa as a medium of instruction (Duka 2018: vi) raises expectations regarding the treatment of academic vocabulary. In the interview, he made special reference to the inclusion of Maths and Science terminology as motivated by his commitment to mother-tongue education (Duka 2021). In order to evaluate the extent of his effort in this respect, curriculum words from the *Oxford Bilingual School Dictionary: IsiXhosa and English* were looked up in the Shuters dictionary. In the Oxford school dictionary, five hundred lemmata, which include instruction words and others used in various school subjects, are marked as important curriculum words. According to the compilers, these important words "have been chosen on the basis of ... frequency of use" in a corpus of "novels, textbooks, official documents" (De Schryver et al. 2015: xi) and other relevant curriculum material. While relying on his intuition and experience as a retired educator and education specialist, Duka (2018) managed to enter seventeen (3.4%) of the Oxford school dictionary's five hundred important curriculum vocabulary. A comparison was also made between the Shuters dictionary's Maths and Science terms and those included in the XNLU's *IsiChazi-magama SeMathematika neNzululwazi* (Maths and Science Dictionary), which is mainly for the intermediate phase. This comparison found that the two dictionaries share about fourteen Maths terms and twenty-four Science terms. The XNLU dictionary, which is a subject-field dictionary, treats more words that are not included in the Shuters dictionary. The comparison of the Shuters dictionary with the other two dictionaries in terms of academic vocabulary coverage might suggest that the envisaged impact of supporting mother-tongue education in isiXhosa might be less than anticipated. Of course, the dictionary also includes a few other words that are not found in the other dictionaries. However, the challenge with estimating its value relates to unclear criteria for including academic vocabulary, including terminology for specific school subjects.

In his quest to provide assistance with respect to contemporary language usage, Duka displays a very strong inclination towards loanwords. This can be seen from Figure 1 below, the first page of the Shuters dictionary.

Aqala ngo-A

aba, ukwaba. (isenzi): Ukwahlula-hlula oko kululutho phakathi kwabantu. Umzekelo: *abantu babelana ngomhlaba ngokulinganayo.*
Izixando: ukwabisa, ukwabeka, ukwabiwa, ukwabela, ukwabelana.
Isibizo: umabi, isabelo.

-abelo, isabelo. *(is-, iz-).* (isibizo): Oko umntu akufumanayo kwinto leyo ekwahlulelwana ngayo. Isenokuba yinto ayizuzayo kwilifa abohlulelana ngalo.
Isininzi: izabelo.

abhakhasi, i-abhakhasi. *(i-).* (isibizo): Isihlohlo sobalo esisetyenziswa kwisifundo seMathematiki.

-abi, umabi. *(um-, ab-).* (isibizo). 1. Umntu owabayo. 2. Umntu onesakhono sokwabela abantu.
Isininzi: ababi.

-adresi, i-adresi *(i-).* (isibizo): 1. Iinkcukacha ezimalunga nendawo umntu ahlala kuyo. 2. Imizekelo: I-adresi yokuhlala, i-adresi yokuthumela imbalelwano okanye i-adresi ye-imeyile.
Isininzi: ii-adresi.

-afidavithi, i-afidavithi. *(i-).* (isibizo): Ingxelo efungelweyo phambi komKhomishina weziFungo.
Isininzi: ii-afidavithi.

-Afrika, i-Afrika. *(i-).* (isibizo): 1. Ilizwekazi elirhanqwe ziilwandlekazi ezintathu ezizezi, i-Atlantiki, iMedithera, nelase-Indiya. 2. Uninzi lwabemi beli lizwekazi ngabantu abamnyama abangabomthonyama.

-Afrika, um-Afrika. *(um-, ama-).* (isibizo): 1. Ummi wezwekazi lase-Afrika ngokwemveli. 2. Kanti ke isenokuba ngumntu wemveli welinye izwekazi (iNtshona, iMpuma, iMelika) othe wamkeleka ngokomthetho wangummi osisigxina e-Afrika.
Isininzi: ama-Afrika.

-Agasti, u-Agasti. *(u-, oo-).* (isibizo): 1. EyeThupha. 2. Ngokwekhalenda yaseNtshona yinyanga yesibhozo enyakeni ukuqalela kweyokuqala enguJanyuwari (eyoMqungu).
Isininzi: oo-Agasti.

-ajenda, i-ajenda. *(i-).* (isibizo): Uludwe lwemicimbi exoxwayo entlanganisweni.
Isihlonipho: *USipho une-ajenda yakhe.* Oku kuthetha ukuba uSipho lo unenye injongo engeyiyo le kudityenwe ngayo. Maxa wambi kuye kuthiwe umntu une-ajenda embaxa. Umntu onjalo uyingozi kuba angakhangeleka ehamba nani kanti ujonge ukunonzakalisa.
Isininzi: ii-ajenda.

akha, ukwakha. (isenzi): 1. Ukunyusa udonga ngezitena okanye ngamatye. 2. Kanti ke usenokwakha umzi ngodaka nezinti. 3. Kukwathetha ukusebenzisa izimvo eziphilileyo zokubumba ubomi bomntu okanye babantu.
Izixando: ukwakhisa, ukwakheka, ukwakhiwa, ukwakhela, ukwakhana.
Isibizo: isakhiwo, umakhi.
Iqhalo: *Intaka yakha ngoboya benye.* Umntu uncedwa ngabanye ukuze aphumelele.

-akhawunti, i-akhawunti. *(i-).* (isibizo): 1. Ityala lomntu elibhalwe ezincwadini kumzi woshishino. 2. Kulindeleke ukuba lo utyalayo ahlawule isikade sakhe.
Isininzi: ii-akhawunti.

Figure 1:   Loanwords on Page 1 of the Shuters dictionary

Six of the twelve (50%) of the lemma on the first page of the Shuters dictionary, namely **abhakhasi** (abacus), -**adresi** (address), -**afidavithi** (affidavit), -**Agasti** (August), -**ajenda** (agenda), -**akhawunti** (account) are loanwords. That the lexicographer prefers loanwords is clearer in cases where they exist alongside indigenous equivalent words. Names of months of the year, such as *August* illustrate this. Indigenous names only appear as part of the microstructural treatment but not as lemmata as can be seen on Figure 1 above. Thus, *eyeThupha* for *August*, *eyoMqungu* (January), *UTshazimpuzi* (April) or *EyoMnga* (December) are not lemmatised in the Shuters dictionary. Lemma selection policies regarding loanwords remain contested in African languages dictionaries. In the case of isiXhosa, the inclusion of loanwords was the major source criticism against the

Oxford Bilingual School Dictionary: IsiXhosa and English (Nkomo 2015).

Besides their inclusion, how loanwords are adapted in relation to the orthographies of African languages remains contentious. Thus, the discussion should not just end by condemning their inclusion, which is not necessarily wrong depending on the functions of dictionaries, among other reasons. For example, the *Oxford English–Xhosa Dictionary*, the *Oxford Bilingual School Dictionary: IsiXhosa and English* and the XNLU monolingual dictionary provide *iadresi* and *idilesi*, as orthographical adaptations of the English word *address*. These spellings are in line with those approved by PanSALB, but the Shuters dictionary has *i-adresi* instead. The Shuters dictionary's departure from existing dictionaries and PanSALB regarding the spelling of loanwords is evident when it renders *April* as *u-Aprili* while the other three dictionaries have *uEpreli* in adherence to the phonetic spelling system. Accordingly, Duka's approach to the inclusion and spelling of loanwords is another area of discussion in the assessment of the dictionary's macrostructure. The remark that was made regarding his non-use of isiXhosa dictionaries that are currently in circulation remains highly relevant as his dictionary tends to deviate, if not regress, regarding matters where existing dictionaries have already made progress.

5.2   The microstructure

This section analyses and evaluates the entries that constitute dictionary articles of the Shuters dictionary. The primary focus will be on items or data items, which are "those entries from which the dictionary user can retrieve some information regarding the subject matter of the … dictionary" (Gouws and Prinsloo 2005a: 116). To determine "the extent of the data categories … included in an article" (Gouws and Prinsloo 2005a: 141), the distinction between obligatory and extended obligatory microstructures is applied in Subsections 5.2.1 and 5.2.2. Thereafter, the discussion will proceed in Subsection 5.2.3 to consider indicators or structural indicators as part of the microstructure of the Shuters dictionary, entries from which the user cannot "retrieve information regarding the subject matter of the dictionary but they … mark a specific item or indicate a specific search field in a dictionary article" (Gouws and Prinsloo 2005a: 116). Altogether, this undertaking evaluates the availability and accessibility of lexicographic information within dictionary entries, and user-friendliness of the dictionary.

5.2.1   The obligatory microstructure

Gouws and Prinsloo (2005a: 141) explain that the obligatory microstructure constitutes of data items included in each and every article. A survey of the Shuters dictionary indicates that all dictionary articles consists of at least the following items (cf. Figure 1 above):

— The lemma sign
— Part of speech data
— Paraphrases of meaning

While each lemma sign is part of the macrostructure of the dictionary, it auto-matically becomes a data item in the obligatory microstructure. The user can retrieve orthographic and morphological information about the lexical items represented by lemma signs, which are entered as stems. Full nouns which show how a stem combines with the relevant prefix follows, e.g. *inceba* imme-diately follows **-nceba** as another data item, while the infinitive form, e.g. *uku*hlela (to edit) immediately follows the lemmatised stem **hlela** (edit). The morphological data is particularly important in the case of nouns whose stems may combine with more than one noun prefix. For example, *ihlelo* (noun class) and *uhlelo* (programme) provided as obligatory entries for two nouns sharing the stem **-hlelo** enables the user to make the necessary morphological distinc-tions that also have semantic and grammatical implications. Consider the case of the stem –**phunga**, which yielded six separate lemmata in the Shuters dic-tionary, five of which are captured in Figure 2 below.

Figure 2:   Five of the six lemmata with the stem *–phunga* in the Shuters dic-tionary

The morphological data that the user derives from the lemma signs, including the indications of stems and prefixes, serves to highlight the productive nature of isiXhosa and Bantu languages in general.

However, the repetition of full nouns and infinitive verbs as compulsory entries is redundant, particularly in the case of verbal lexical items since there are few instances where the variant prefix *ukw-* is used instead of *uku-*. The same effect regarding orthographical and morphological guidance could be achieved with an added benefit of saving space. This would entail simply including the prefixes in brackets after the lemma sign or attaching the prefix on the left of the lemma sign and excluding the full forms given in the Shuters dictionary. Verbal lexical items could be presented as follows:

> **Hlela** (uku-) … or uku**Hlela** …
> **Ela** (ukw-) … or ukw**Ela** …

Both these approaches are notable in other isiXhosa and Nguni dictionaries. The XNLU's *IsiChazi-magama SesiXhosa* is a good example of a dictionary that exhibits the latter approach (see Figure 3 below). As the guiding elements regarding the location of the lexical items belonging to different grammatical categories, stems appear in bold face with relevant prefixes, which are in regular font, attached to them on the left. The dictionary, thus, displays a left-expanded article structure, a notion introduced by Gouws and Prinsloo (2005b). Had this approach been applied in the Shuters dictionary, the related isiXhosa terms for *noun class* and *edition*, which share *-hlelo* as the stem could be entered as i**Hlelo** and u**Hlelo** respectively.

The relevance of part of speech entries in dictionaries is a point that will not be belaboured in this article. Part of speech information can assist users with guidance for sense discrimination in cases where certain words belong to different parts of speech. Gouws and Prinsloo (2005a: 125-126) offer useful examples of such information for both text reception and text production purposes. The Shuters dictionary highlights many instances where the same stem results in multiple words belonging in different grammatical categories, e.g. *-phunga* as shown in Figure 2 above. Although the hyphen distinguishes between noun and verbal lemmata, some inexperienced dictionary users might have to rely on part of speech data to confirm if they are looking at the right lemma.

For the Shuters dictionary, meaning information is alluded to as the most important when the lexicographer makes an emphatic statement about the language's capability to explain isiXhosa words (Duka 2018: v), perhaps in contrast to other isiXhosa dictionaries which provide meaning through English equivalents. It is important to note that this dictionary provides meaning information primarily by means of concise paraphrases of meaning. For example, the meaning of **nyamezela** (persevere) is explained as "Ukumelana neenzima neembandezelo neminye imingeni unganikezeli" (To withstand difficulties, oppression and other challenges without giving up). An equally informative

but wordy paraphrase of meaning is provided in the XNLU monolingual dictionary as shown below.
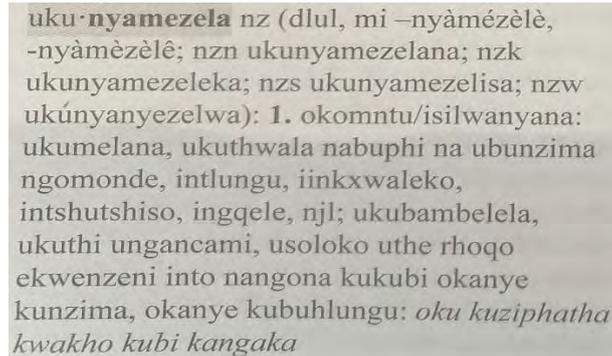
Figure 3:   The explanation of the meaning of *-nyamezela* in the XNLU monolingual dictionary

No real additional information may be obtained from the above article. Instead, some potentially more difficult words and expressions such as *iinkxwaleko*, *intshutshiso* and *uthe rhoqo*, may compel further searches, thereby burdening the user and distracting them from the activity that prompts the initial dictionary consultation. Thus, the concise paraphrases of meaning as the default data for explaining meaning in the Shuters dictionary is a massive improvement for mother-tongue speakers of isiXhosa who had to rely on mainly translation equivalents in the bi- and multilingual dictionaries. However, synonym definitions are also used needlessly in some cases in the Shuters dictionary. Consider the following example illustrating the provision of meaning information for the word **umbuzo** (question):
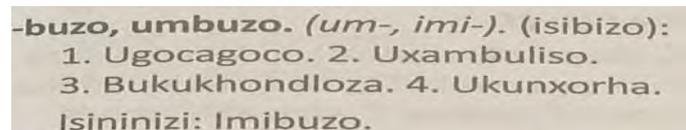
Figure 4:   Explanation of *umbuzo* using synonyms in the Shuters dictionary

For the user to benefit from such an article, knowing the meaning of synonyms such as *ugocagoco, uxambuliso, bukukhondloza* and *ukunxorha* is imperative. Such knowledge may not be taken for granted among young speakers of the language, especially given the diminished space for African languages in the South African education system and the permeation of English even in the

domestic spheres. Young language users no longer have adequate exposure and access to rich vocabulary in their languages.

The treatment of polysemy is another aspect for which the Shuters dictionary might be commended. The lexicographer has provided numbered paraphrases of meaning for lexical items with multiple senses. However, unnecessary repetition of the same senses with slightly different formulations undermines this effort. The screenshot in Figure 5 below is a case in point with respect to the lemma **um**a**khi**.
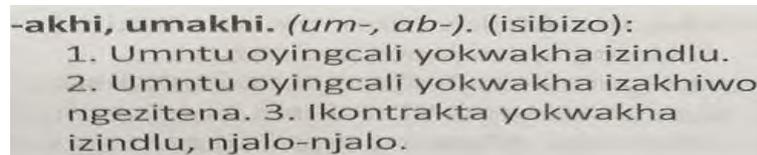
> **-akhi, umakhi.** *(um–, ab–).* (isibizo):
> 1. Umntu oyingcali yokwakha izindlu.
> 2. Umntu oyingcali yokwakha izakhiwo ngezitena. 3. Ikontrakta yokwakha izindlu, njalo-njalo.

Figure 5:   Article for the lemma *umakhi* in the Shuters dictionary

The first sense translates into "A person who is an expert in building houses". The second one would read as "A person who is an expert in constructing buildings using bricks" in English. The two paraphrases of meaning do not describe different senses of the word but differ in detail. Such cases are prevalent and they undermine the effort of capturing multiple senses of lexical items in the Shuters dictionary. At the same time, this practice takes up more space without providing additional help for the users of the dictionary.

5.2.2   Extended obligatory microstructure

In addition to default entries such as those identified in 5.2.1 above, extended obligatory microstructures refer to dictionary articles with extra data, depending on the types of lexical items represented by lemma signs, their parts of speech and semantic relations (Gouws and Prinsloo 2005a: 142). Nouns and verbs in the Shuters dictionary display an extended obligatory microstructure.

5.2.2.1 Additional data for nouns

For all lemmata falling under the part of speech category of nouns, prefixes are indicated for both singular and plural forms in the Shuters dictionary. These are respectively represented by 'sing. prefix' and 'pl. prefix' in the second and third columns of Table 1 below. However, the explicit class indication given in brackets is not part of the dictionary articles and possibly a regrettable omission for the potential dictionary users. The first column provides examples of nominal lemmata, which the lexicographer lemmatised according to stems

before giving the full singular forms. The obligatory data given in the first three columns of Table 1 constitutes morphological information that could be valuable in the study of grammar at high school and university levels.

Table 1:   Morphological data for nominal lexical items in the Shuters dictionary

Lemma	Sing. Prefix	Pl. Prefix	Plural Form
-akhi, umakhi (builder)	um- (class 1)	ab- (class 2)	abakhi (builders)
-mama, umama (mothers)	u- (class 1a)	oo- (class 2a)	oomama (mothers)
-lambo, umlambo (river)	um- (class 3)	imi- (class 4)	imilambo (rivers)
-langa, ilanga (sun/day)	il- (class 5)	ama- (class 6)	amalanga (-/days)
-iselo, isiselo (drink)	isi- (class 7)	izi- (class 8)	iziselo (drinks)
-ja, inja (dog)	in- (class 9)	izin- (class 10)	izinja (dogs)
-thi, uluthi (stick)	ulu- (class 11)	izin- (class 10)	izinti (sticks)
-bomi, ubomi (life)	ubu- class 14)	—	
-tya, ukutya (food)	uku- (class 15)	—	

Apart from the indication of singular and plural prefixes, the extended obligatory microstructure for nominal lemmata includes additional data that apply only to certain types of lexical items. The first type of data is the indication of full plural forms, showing how noun stems combine with different plural prefixes. Examples are given in column 4 of Table 1 above. However, this data type does not constitute part of the extended obligatory microstructure of nominal lemmata in general since not all nouns included in the Shuters dictionary are countable. Nouns denoting uncountable phenomena such as *umunyu* (sympathy) and *amandla* (power/energy) do not have plural forms included as part of dictionary articles. Although sympathy and power or energy may be quantifiable in general terms, they are nevertheless uncountable. However, the first one belongs in class 3 and morphologically looks like countable singular nouns belonging in that class, while the latter belongs in class 6 together with plural forms of nouns belonging in class 5.

In isiXhosa and other Bantu languages, suffixes are added to create forms that bear certain semantic dimensions such as the female gender, augmentation and diminutives. Table 2 below illustrates some examples from the Shuters dictionary.

Table 2:   Nominal suffixing in the Shuters dictionary

Feminine gender	Augmentation	Diminutives
inja > inja*kazi*	isixeko > isixeko*kazi*	inkomo > inkony*ana*
igqirha > igqirha*kazi*	impixano > impixano*kazi*	umntu > umntw*ana*
Itshawe > itshawe*kazi*	ipoma > ipoma*kazi*	indlu > indlw*ana*
umqingqi > umqingqi*kazi*		icephe > icetsh*ana*
umprofeti > umprofeti*kazi*		Ithumbu > ithunj*ana*

The suffix *-kazi* is used is used as a gender marker and for augmentation. As a gender marker, it denotes that the subjects referred to in the first column of Table 2 are biologically female. No social connotations or constructions of **inja***kazi* (female dog) or **umprofeti***kazi* (prophetess) being weaker than the male counterparts are expressed by the suffix. However, when used in the cases of lemmata in the second column, the *-kazi* suffix augments the subject/object to which the noun refers. Thus, **isixeko***kazi* is a big city while **impixano***kazi* is a fierce conflict. In the case of personal names such as *Siphokazi*, *Vuyokazi* and *Thandokazi* given to females, both the gender and augment senses would appear to be in complementary use. Firstly, they seems to acknowledge that the individuals are females, while at the same time indicating that the phenomena after which they are named, e.g. *sipho* (gift), *vuyo* (joy) and *uthando* (love), are huge or abundant. This has generated contestations of whether African languages such as isiXhosa are gendered or not, with some scholars attributing *-kazi* as a gender marker to colonial linguistics (Maseko 2018). However, a deep engagement with those debates is beyond the scope of this article.

As a diminutive suffix, *-ana* carries a sense of young, little, small or tiny, depending on the context in which nouns are used in Bantu languages such as isiXhosa. The Shuters dictionary included diminutive forms such as those given in the third column of Table 2 as optional data at the end of some articles whose lemma signs are nouns. In the case of **umntu** (person) and **inkomo** (cow), *umntwana* (child) and *inkonyana* (calf) as diminutive forms denote the young ones, while indicating that the house, spoon and intestine are of small sizes in *indlwana*, *icetshana* and *ithunjana* respectively.

Nominal lemmata marked for gender, augmentative and diminutive forms constituting the extended obligatory microstructure of the Shuters dictionary provide morphological and semantic data that could be useful for dictionary users. However, these forms are indicated sparingly in the dictionary. What is not clear is the criteria that was used to decide on the inclusion of those forms that are entered. The augmentative and diminutive suffixes combine with much more nouns than was indicated in the dictionary.

5.2.2.2 Additional data for verbs

Verbal extensions constitute key microstructural entries for all verbal lemmata in the Shuters dictionary. Many scholars acknowledge the prominence of verbal extensions in Bantu languages such as isiXhosa (Bosch, Pretorius and Jones 2007; Cocchi (2009); Nurse and Philippson 2003). Cocchi (2009: 93) observes "several important differences concerning function, productivity, combinatory possibilities and mutual exclusion patterns" of verbal extensions. These differences make the inclusion of verbal extensions an important undertaking, especially for dictionaries that are used in the teaching/learning of grammar for productive purposes. However, the combinatory possibilities and exclusion patterns need to be taken into account. According to Bosch et al. (2007: 135), "verbal extensions are not able to combine with all verb roots". This "necessitates the explicit inclusion of known occurrences as subentries under the base form" (Bosch et al. 2007: 135). Accordingly, while verbal extensions feature for all verbs in the Shuters dictionary, some extensions do not appear for certain verbs. As an example, the reciprocal extension would not apply in verbs whose possible objects do not possess any agentive potential. Thus, for the verb **cula** (sing), *ukuculisa* (to cause to sing), *ukuculiseka* (to be capable to be caused to sing), *ukuculwa* (to be sung), *ukuculela* (to sing for), and *ukuculisana* (to cause each other to sing) are included while *ukuculana* (to sing each other) is excluded. However, the omission of *ukuculeka* (to be capable to be sung) is questionable.

Related to the inclusion of verbal extensions, pseudo-extended verb stems pose another tricky challenge for lexicographers. Phiri (1980: 63) defines a pseudo-extended verbal stem as "that minimal stem whose final element resembles a verb extension which cannot be isolated". He adds that "[t]he resemblance to an extension is usually reinforced by a characteristic meaning common to the series of such stems" (Phiri 1980: 63). Verbs with such stems are abundant in isiXhosa. They require lexicographers to be meticulous and avoid treating them as extended verbs. Examples include **landela** (follow) **phangela** (work) and **thelekisa** (compare) which are included as lemmata in the Shuters dictionary. Their stems could erroneously be treated as extended stems of *landa* (collect) *phanga* (to loot or take forcefully) and *thela* (pour). Yet they do not display strong semantic connections with the former. By including them as lemmata in their own right, the lexicographer avoided such confusion. That he further included applicable extended forms for such lemmata, e.g. *ukuphangelisa* (to cause to work), *ukuphangeleka* (to be able to make work possible), *ukuphangelwa* (to be worked), *ukuphangelela* (to work for), *ukuphangelisana* (to cause each other to work) is commendable. This affords the lemma **phangela** the same treatment as **phanga**. However, inconsistency is noted once again, with verbal extensions not featuring under the article for **landela**.

5.2.2.3 Proverbs and idioms

Proverbs and idioms constitute the final types of data items provided as part of

some dictionary articles in the Shuters dictionary. Including and explaining them is useful for text production and text reception They are included as part of the extended obligatory microstructure for words serving as bases for the proverbs or idioms, indicated in isiXhosa as *isaci* and *iqhalo* respectively. The article for **khotha** illustrates this:

khotha, ukukhotha. (isenzi): Ukulenya.
Umzekelo: *inkomo ikhotha enye ngolwimi.*
Izixando: ukukhothisa, ukukhotheka, ukukhothwa, ukukhothela, ukukhothana.
Isaci: *Ukukhoth' exathula.* Uncoma egxeka.
Iqhalo: *Inkomo ikhoth' eyikhothayo.* Umntu unceda omncedayo.
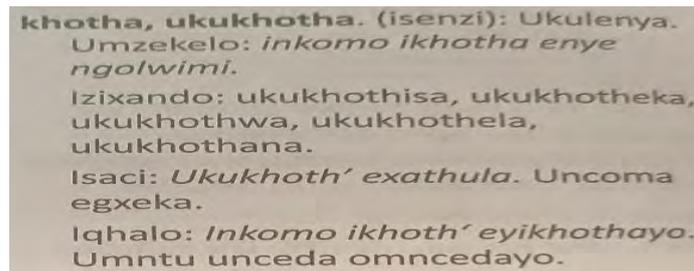
Figure 6:   Article indicating an idiom (isaci) and a proverb (iqhalo) based on *khotha*

The lemma **khotha** (lick/caress) serves as a base for an important proverb and idiom isiXhosa. The proverb refers to the habitual and reciprocal act of cattle which caress each other using tongues, i.e. *Inkomo ikhoth' eyikhothayo.* Literally, this proverb can be glossed as 'a cow caresses another cow that also caresses it', meaning that a person helps those who also help them). The idiom *ukukhoth' exathula* literally means that one caresses while scratching, which would refer to a diplomatic or treacherous act of criticising while also complimenting.

While the approach taken by Duka is common in many dictionaries, it presents challenges when there is inadequate certainty about the base among the different words constituting a proverb or an idiom. The idiom *imbiza ibona iketile* is a good example. Literally, it means 'the pot sees the kettle'. It refers to someone who sees faults other people's faults but not their own. The lexicographer included it under both **imbiza** (pot) and **iketile** (kettle). If the user is to benefit from proverbs and idioms, they need know their bases. If the user does not know, they may have to access these data types fortuitously under the relevant lexical items even if the need for the data had not initiated the dictionary consultation process.

5.2.3   Structural indicators

Gouws and Prinsloo (2005a: 116) state that structural indicators "are not entries from which the user can retrieve information regarding the subject matter of the dictionary but they … mark a specific item or indicate a specific search field in a dictionary article". Their distinction between typographical and non-typo-

graphical structural indicators applies to all print dictionaries including the Shuters dictionary. The different screenshots taken from the Shuters dictionary highlight three typographical indicators, namely bold print that is used for lemma signs, italics for examples and regular Roman for the rest of the data types. Non-typographical indicators complement the typographical ones to offer guidance with respect to the identification of data items and indication of specific search fields within the dictionary articles. The following non-typographical indicators are identified with their functions:

(1)   **Punctuation marks**
    — commas: (a) separating the lemma sign and the full verb in infinitive form and (b) separating different verbal extensions
    — full stops: (a) after the infinitive verb, (b) at the end of each paraphrase of meaning, (c) after the idiom; and (d) at the end of the article
    — brackets in which part of speech data is provided
    — colon: (a) before the first paraphrase of meaning, (b) before the example, (c) verbal extensions; and (d) before the idiom.

(2)   **Linguistic labels**
    — isenzi (verb)/isibizo (noun) for indicating part of speech
    — Umzekelo (example)
    — Izixando (verbal extensions)
    — Isaci (idiom)
    — Iqhalo (proverb)

(3)   **Numerals**
    — Used for sense discrimination

The most remarkable feature of the Shuters dictionary is the avoidance of abbreviations and place keeping symbols to indicate parts of speech and other data types. The lexicographer opted for full linguistic labels such as *isibizo* (noun), *isenzi* (verb), *isininzi* (plural) *umzekelo* (example), *isaci* (idiom), *iqhalo* (proverb), *isikhomokazi* (female gender marker), etc. These non-typographical indicators are used explicitly even in cases where implicit indication could be used. For instance, the indication of examples could have been done implicitly using italics only, as done in many print dictionaries. All this indicates that the lexicographer intended to be as explicit as possible in order to make the dictionary accessible. This is a considerate decision, given that school learners appear to be the primary target users of the dictionary. The unfortunate part is that all these structural indicators are not explained anywhere in the dictionary.

5.3   Outer texts

The analysis of the Shuters dictionary has thus far focused on the contents of

dictionary articles. Gouws (2004: 68) is critical of "[t]he traditional bias towards the contents of dictionary articles as the main field of interest for dictionary users, lexicographers and metalexicographers". He notes that such an approach is "detrimental to the study, the development and the use of the structure of a dictionary and the different structural components" (Gouws 2004: 68). The main-text bias by lexicographers limits the potential of the dictionary in terms of data distribution and the realisation of its functions. On the part of the user, it results in the under-utilisation of lexicographic data that is included in the outer texts. In an effort of making this exercise of dictionary criticism to be as judicious as possible, it is important to consider its outer texts of Shuters dictionary.

The study of outer texts recognises dictionaries as compound texts and it has led to useful characterisations taking into account the relationship between outer texts and main texts in the accomplishment of dictionary functions (cf. Bergenholtz, Tarp and Wiegand 1999; Gouws 2004; 2007b; Kammerer and Wiegand 1998). Drawing from this existing literature, a partially extended frame structure (cf. Gouws 2004: 68-9) best characterises the Shuters dictionary. The dictionary contains only the front matter, but no back-matter texts to make it completely extended. Apart from the title page, imprint and the acknowledgement text, which may be classified as non-integrated outer texts (Gouws 2007b), the other texts may be categorised as integrated texts as they play a role in accomplishing at least one of the dictionary's lexicographic functions (Gouws 2007b). The latter include the table of contents, which guides the user to other front matter texts and the beginning of individual alphabetic lemma stretches in the main text. The other texts were conceptualised as integrated function-adhering outer texts, as the lexicographer writes:

> Intshayelelo le yenziweyo yeyokuncedisa umfundi aqonde ezinye zezinto aza kuhlangana nazo kulo mqulu. Kanti kanjalo ikwancedisa utitshala ekuxhobiseni umfundi ekuqondeni isiXhosa ngokubanzi nanzulu.

> The introduction that has been given is to assist the learner to understand some of the things that s/he will encounter in this compilation. At the same time, it will help the teacher in equipping the learner with understanding isiXhosa broadly and deeply (Duka 2018: xvi).

The statement applies to the front matter texts dealing with various aspects of the language that appear in the main text. These include grammatical texts alluded to earlier, which describe some phonological, morphological and syntactical aspects of isiXhosa, the treatment of loanwords, idioms and proverbs. To a certain extent, these texts also serve as a user guide with respect to specific linguistic phenomena that fall within the lexicographic treatment programme of the dictionary. This is made explicit in some of the texts. For example, the texts dealing with loanwords, compound nouns and verbal extensions are excellent in that respect. However, what is regrettable is the brevity of the available texts and the absence of a user guide which deals with aspects of dic-

tionary structures. The grammar section could have been made more comprehensive and systematic as well in order to enhance the cognitive and communicative functions that the lexicographer alludes to in the introductory and blurb texts.

6.   Conclusion

This article undertook a critical evaluation of the *Shuters IsiChazi-magama SesiXhosa* as a recent addition to the inventory of isiXhosa dictionaries. Apart from being relatively new, what makes this dictionary special is that it is only the second monolingual dictionary in the language, as outlined in the introduction. Therefore, the main endeavor of the article was to determine the contribution that this dictionary makes to isiXhosa lexicography in view of its predecessors in the language, including the pioneer monolingual dictionary. The undertaking drew theoretical and methodological insights from dictionary criticism as a sub-field of lexicographic research. The article focused on the macrostructure, microstructure and the front matter of the dictionary.

When viewed against *IsiChazi-magama SesiXhosa* compiled by the XNLU ten years earlier, the Shuters dictionary may be characterised as a basic monolingual dictionary. Of its envisaged target users, the school market may benefit more than university students and the generality of isiXhosa speakers. Concise paraphrases of meaning compete favourably with those of the XNLU's *IsiChazi-magama SesiXhosa*. However, its excessive use of synonyms reduces the user-friendliness of the dictionary as users may be forced onto extended and circular searches for meaning information. Needless repetition of senses may also mislead users into thinking that many lexical items treated in the dictionary are polysemous. Furthermore, the macrostructural composition of the dictionary exhibits the lexicographer's intuitive approach to lemma selection, which resulted in an unbalanced representation of vocabulary belonging from different grammatical categories and registers used in different school subjects for the primary target users of the dictionary. The adaptation of loanwords could also shorten the dictionary's lifespan in the market, as it departs from some orthographic rules established by PanSALB and embraced in existing dictionaries. Finally, better utilisation of the outer texts to enhance the data distribution and purposeful integration with the main text could have resulted in a more user-friendly dictionary. Nevertheless, another dictionary in an African language produced without the involvement of a National Lexicography Unit is welcome towards further intellectualisation of African languages. What is needed is a better application of insights from metalexicography.

Endnote

1.   The Duka (2021) reference is based on a telephone interview between this author and M.M.M. Duka regarding the dictionary, his motivation, his approach to lexicographic task and relevant expertise that guided his work.

Acknowledgement

References

Alberts, M. 2011. National Lexicography Units: Past, Present, Future. *Lexikos* 21: 23-52.

Awak, M.K. 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989*: 8-18. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-language and Terminology Research*: 1762-1832. Berlin/New York: Walter de Gruyter.

Bongela, K.S. 1991. *Amagontsi*. Umtata: Afro Publishers.

Bosch, S.E., L. Pretorius and J. Jones. 2007. Towards Machine-readable Lexicons for South African Bantu Languages. *Nordic Journal of African Studies* 16(2): 131-145.

Busane, M. 1990. Lexicography in Central Africa: The User-Perspective, with Special Reference to Zaïre. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989*: 19-35. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.

Cocchi, G. 2009. Bantu Verbal Extensions: A Cartographic Approach. Moscati, V. and E. Servidio (Eds.). 2009. *Proceedings XXXV Incontru di Grammatica Generativa. Vol.3:* 92-104. Siena: Centro Interdiparmentiale di Studi Cognitivi sul Linguaggio.

De Schryver, G.-M. 2008. A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary. *Lexikos* 18: 63-91.

De Schryver, G.-M. and A. Wilkes. 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and D. DeCesaris (Eds.). 2008. 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 827-836. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

De Schryver, G.-M. et al. (Eds.). 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English*. Cape Town: Oxford University Press Southern Africa.

Doke, C.M. and B.W. Vilakazi. 1948. *Zulu–English Dictionary*. Johannesburg: Witwatersrand University Press.

Duka, M.M.M. 2018. *Shuters IsiChazi-magama SesiXhosa*. Pietermaritzburg: Shuter and Shooter.

Duka, M.M.M. 2021. Interview, 28 June 2021.

Duka, M.M.M., S.W. Mhlontlo and T. Matoto. 2010. *Isihobe se-Afrika Entsha*. Pietermaritzburg: Shuter and Shooter.

Fischer, A., E. Weiss, E. Mdala and S. Tshabe. 1985. *Oxford English–Xhosa Dictionary.* Cape Town: Oxford University Press Southern Africa.

Gallardo, A. 1980. *Dictionaries and the Standardisation Process*. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia, South Carolina: Hornbeam Press.

Gouws, R.H. 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.

Gouws, R.H. 2007a. On the Development of Bilingual Dictionaries in South Africa: Aspects of Dictionary Culture and Government Policy. *International Journal of Lexicography* 20(3): 313-327.

Gouws, R.H. 2007b. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.

Gouws, R.H. 2017. Dictionary Criticism and Dictionary Critics. Bielińska, M. and S.J. Schierholz (Eds.). 2017. *Wörterbuchkritik — Dictionary Criticism*: 33-49. Berlin/Munich/Boston: De Gruyter.

Gouws, R.H. and F.A. Ponelis. 1992. The Development of Afrikaans and the Lexicographical Tradition. Zgusta, L. (Ed.). 1992. *History, Languages and Lexicographers*: 77-104. Tübingen: Max Niemeyer.

Gouws, R.H. and D.J. Prinsloo. 2005a. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Gouws, R.H. and D.J. Prinsloo. 2005b. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London: Routledge.

IsiXhosa National Lexicography Unit. 2013. *Isichazi-magama seMathematika neNzululwazi*. Cape Town: South African Heritage Publishers.

Kammerer, M. and H.E. Wiegand. 1998. Über die textuelle Rahmenstruktur von Wörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.

Kaschula, R.H. and D. Nkomo. 2019. Intellectualisation of African Languages: Past, Present and Future. Wolff, H.E. (Ed.). 2019. *The Cambridge Handbook of African Linguistics*: 601-622. Cambridge: Cambridge University Press.

Kropf, A. 1899. *A Kafir–English Dictionary*. Alice: Lovedale Press.

Maseko, P. 2018. Language as Source of Revitalisation and Reclamation of Indigenous Epistemologies: Contesting Assumptions and Re-imagining Women Identities in (African) Xhosa Society. Bam, J., L. Ntsebeza and A. Zinn (Eds.). 2018. *Whose History Counts: Decolonising African Precolonial Historiography*: 35-55. Stellenbosch: SUN PReSS.

McLaren, J. 1915. *A Concise Kafir–English Dictionary*. London: Longmans, Green and Company.

Meinhof, C. and N.J. van Warmelo. 1932. *Introduction to the Phonology of the Bantu Languages*. Berlin: Reimer/Vohsen.

Mugglestone, L. 2015. *Samuel Johnson and the Journey into Words*. Oxford/New York: Oxford University Press.

Nielsen, S. 2009. Reviewing Printed and Electronic Dictionaries: A Theoretical and Practical Framework. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century*: 23-41. Amsterdam: John Benjamins.

Nkomo, D. 2015. From Common Conversations around the *Oxford Bilingual School Dictionary: IsiXhosa–English* to Constructive Dictionary Criticism. Bosch, S.E. and D. Nkomo (Eds.). 2015. *Programme and Abstracts: 20th International Conference of the African Association for Lexicography, 6–8 July 2015, University of KwaZulu-Natal, Durban, South Africa.*

Nkomo, D. 2018. Dictionaries and Language Policy. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 152-165. New York: Routledge.

Nkomo, D. 2020. Dictionary Culture in African Language Communities: Research, Development, Challenges and Prospects. *Lexicographica* 36(1): 11-37.

Nurse, D. and G. Philippson. 2003. Introduction. Nurse, D. and G. Philippson (Eds.). 2003. *The Bantu Languages*: 1-12. London/New York: Routledge.

Oosthuysen, J.C. 2016. *The Grammar of isiXhosa*. Stellenbosch: SUN PReSS.

Pahl, H.W. et al. 1967. *IsiXhosa seBanga leMatriki.* Johannesburg: Educum Publishers.

Phiri, A.G. 1980. *A Description of Derivational Forms in Tumbuka*. Doctoral dissertation. Austin: The University of Texas.

Sigcu, A.T. and L.T. Manyase. 1981. *Nasi isiXhosa: Ibanga* 7. Umtata: Shuter and Shooter.

Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.

Swanepoel, P. 2017a. The Contribution of Dictionary Criticism to Dictionary Research. Bielińska, M. and S.J. Schierholz (Eds.). 2017. *Wörterbuchkritik — Dictionary Criticism*: 11-31. Berlin/Munich/Boston: De Gruyter.

Swanepoel, P. 2017b. Methods in Dictionary Criticism. Bielińska, M. and S.J. Schierholz (Eds.). 2017. *Wörterbuchkritik — Dictionary Criticism*: 81-112. Berlin/Munich/Boston: De Gruyter.

Tarp, S. 2017. Dictionary Criticism and Lexicographical Function Theory. Bielińska, M. and S.J. Schierholz (Eds.). 2017. *Wörterbuchkritik — Dictionary Criticism*: 113-132. Berlin/Munich/Boston: De Gruyter.

Tshabe, S.L., Z. Guzana and A.B.B. Nokele. 2008. *IsiChazi-magama SesiXhosa*. Pietermaritzburg: Nutrend Publishers.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

The Real-life-based
School Dictionary for Turkish*

Bülent Özkan, *Turkish Language Teaching Department,*
Mersin University, Turkey (ozkanbulent@mersin.edu.tr)

Abstract: This report aims to describe the development of a new, original, and monolingual dictionary as an educational resource named the Real-life-based School Dictionary (RLBSD). There is a general lack of corpus-based school dictionaries in the Turkish context. The existing dictionaries are based on native speaker intuitions and are far from being empirical, indicating the need for the development of a corpus-based dictionary. In this respect, the RLBSD meets this critical need. A specialized corpus consisting of texts from a range of natural language environments such as coursebooks (from primary to high school) approved by the Ministry of Education (MoNE), children's literature books, and periodicals was compiled and forms the basis of the RLBSD. The dictionary has been designed with the use of information technologies and corpus linguistics methods and is accessible to users on all digital platforms (the Internet, smart board, tablet, smartphone) as an instructional tool. It is believed that the current study may serve as an exemplary work with its methods and outputs for Turkish lexicography studies.

Keywords: SCHOOL DICTIONARY, LEXICOLOGY, TURKISH EDUCATION, REAL-LIFE-BASED DICTIONARY, CORPUS LINGUISTICS

Opsomming: Die werklikheidsgebaseerde skoolwoordeboek vir Turks. In hierdie verslag word die ontwikkeling van 'n nuwe, oorspronklike, en eentalige woordeboek as opvoedkundige hulpbron, naamlik die Real-life-based School Dictionary (RLBSD), beskryf. Daar bestaan 'n algemene tekort aan korpusgebaseerde skoolwoordeboeke in die Turkse konteks. Die bestaande woordeboeke is gebaseer op die intuïsies van moedertaalsprekers en is nie naastenby empiries gefundeer nie, wat dui op die behoefte aan die ontwikkeling van 'n korpusgebaseerde woordeboek. Ten opsigte hiervan, voorsien die RLBSD in hierdie kritiese behoefte. 'n Gespesialiseerde korpus, bestaande uit 'n groot verskeidenheid tekste uit natuurlike taalomgewings soos handboeke (van primêre tot hoërskool) wat deur die Ministerie van Opvoedkunde (MoNE) goedgekeur is, kinderlektuur, en tydskrifte is saamgestel en vorm die basis van die RLBSD. Die woordeboek is ontwerp met gebruikmaking van inligtingstegnologie en korpuslinguistiese metodes en is

---

op alle digitale platforms (die Internet, interaktiewe witbord, tablet, slimfoon) vir gebruikers toe-ganklik as onderrighulpmiddel. Daar word aanvaar dat die huidige studie met sy metodes en resultate as voorbeeld kan dien vir Turkse leksikografiese studies.

Sleutelwoorde: SKOOLWOORDEBOEK, LEKSIKOLOGIE, TURKSE OPVOEDKUNDE, WERKLIKHEIDSGEBASEERDE WOORDEBOEK, KORPUSLINGUISTIEK

1.   Introduction

The terms *school dictionary*, *children's dictionary,* and *college dictionary* in lexico-graphic literature all refer to "a type of dictionary written for school-age chil-dren" (Hartmann and James 1998: 122) and have certain common characteristics. They contain a controlled vocabulary with a simple design often supported with visuals. The lemma list in this type of dictionary requires systematic selection.

In the development stage of a dictionary, certain issues such as *readability or usability* (readership), *dictionary format*, *content*, *lexical item presentation* (e.g., part of speech, derivational/inflectional affixes and dependent morphemes to be provided or not), the stages of compiling the dictionary and its publication should be considered (Bowern 2008). Similarly, in developing a school diction-ary, these issues should also be taken into account.

User research in lexicography has gained much importance in the devel-opment of dictionaries. Therefore, what users expect from a dictionary and what would be useful to include in it are among the common practices referred to in the literature. In this respect, the determination of the dictionary's target users is crucial (Jackson 2002). Various field studies such as surveys, observa-tions, or expert opinions are applied to determine the target users of a diction-ary (Atkins and Rundell 2008: 30). The target users, their age range, level of education, and intended purpose of the dictionary have profound effects on the whole structure of a dictionary.

There is a consensus in the literature on the necessity of the target popula-tion's needs in the development process of a dictionary. For this reason, needs-analysis using empirical methods should be performed to determine the needs of the target users (Atkins and Levin 1995: 85). In the current project, the needs analysis was conducted to determine the needs of the target dictionary users. Moreover, in the Turkish context, ineffective lexicographical traditions such as using index cards, finding example sentences manually, basing headwords, and example selection on intuitions are common practices in the development of a Turkish dictionary. However, in the development of a dictionary, the use of corpus is becoming the mainstream as it eliminates these ineffective lexico-graphical traditions. For this reason, RLBSD in the current project was devel-oped with the use of corpus.

Individual preferences and intuitive approaches of linguists and lexicographers necessarily yield different results (McEnery et al. 2006: 145). With an approach based on empirical results and language use research, modern lexicography develops dictionaries through language databases or corpora representing the language they are compiled for, yielding significant results for both lexicographers and users.

Along with the developments in corpus linguistics and corpus tools and their effect on lexicography today, there are few dictionary practices not based on corpus data. In the Turkish context, however, there has been a profound lack of corpus-based dictionaries. Dictionaries on Turkish should be developed based on current lexicographical practices including the development of dictionaries with corpus-based approaches. In this regard, the development of RLBSD as described in this report can be considered as a starting point to fill this important gap.

Lexicography is, by its nature, a discipline that offers empirical and scientific ways for determining vocabulary also including the phases such as specifying lexical entries and providing definitions and examples for a dictionary. As in large-scale dictionaries, school dictionaries' compilation requires considering certain issues such as consistency, being free of discrepancies, and testability of lemma selection by the experts. School dictionaries should be constantly developed to be better by correcting the mistakes and making sure that there is harmony with interrelated, appropriate, and reasonably organized entries. If lemma selection in a dictionary is not based on scientific criteria, it cannot fully represent the vocabulary of the target language. Therefore, it is of great importance to choose lexical items to be included in a school dictionary from the lexical items of the target language vocabulary (Bozkurt 2017).

Research on Turkish school dictionaries indicates that the need to use a dictionary increases with age and school grade in middle schools due to the increasing intensity of the course contents and high vocabulary variation in the texts (Melanlıoğlu 2013). Akcan (2012) examined Turkish workbook exercises and found that students are directed to use dictionaries for 123, 174, 183, 186, 167, and 853 words in the first, second, third, fourth, fifth, and sixth grades, respectively. The same study emphasized that dictionaries should be used as a resource for word meaning and form and usage (Akcan 2012).

The studies on existing school dictionaries report that students, as dictionary users, mention problems such as lack of bold letters for word entries, examples for connotations and figurative meanings, and visual elements (Melanlıoğlu 2013). As is observed in the literature, school dictionaries are developed without initially-set criteria, a reasonable explanation for the visuals to be included that accompany words principles to be used for limiting the content of lexical entries (Çotuksöken 1999).

The current project set out to meet the user group's needs by identifying the vocabulary used in natural language environments that children encounter, evaluating the effectiveness of a dictionary in educational settings, and calculating fit indices based on expert opinion. The RLBSD will contribute to the literature as an original and contemporary work in terms of its processes for development, expansion/distribution, and evaluation, and being a resource based on empirical results. In this regard, an original and real-life-based school dictionary requires a *lexicographical* practice featuring *corpus linguistics* principles and methods and also a special purpose *corpus* containing texts from natural language environments such as coursebooks, children's literary books, and periodicals that children may encounter as an educational resource for the primary and elementary school students.

1.1   Aim and objectives

This report aims to describe the development of a new, original, and monolingual Real-life-based School Dictionary (RLBSD) as educational material. To this end, two objectives were sought to be achieved: the compilation of a special purpose *corpus* containing texts from natural language environments such as MoNE (Ministry of National Education)-approved and current coursebooks, children's literary books, and periodicals that children may encounter starting from elementary school to high school, and development of the Real-life-based School Dictionary based on the corpus in question.

This project's primary objective is to develop a real-life-based school dictionary as educational material by using information technologies and corpus linguistics methods. In the literature, there is a lack of a corpus-based school dictionary in Turkish lexicology studies. Moreover, existing dictionaries are based on native speaker intuitions making these dictionaries far from being empirical. In this respect, the RLBSD as an educational material meets a critical need that has been felt so far. The vocabulary that belongs to the natural language environments that school-age children encounter was determined through lexicographical practices.

1.2   Compilation of the corpus

The research corpus was developed using corpus linguistics principles and methods as a discipline that can yield real-life and empirical outcomes based on texts from natural use environments, potentially providing the researchers the most suitable and up-to-date opportunities for creating a new, original, and real-life-based school dictionary.

The MoNE-approved coursebooks and instructional materials were included in the specialized corpus to reveal their vocabulary. Other natural lan-

guage environments (e.g., children's literary books, periodicals, etc.) were demonstrated in a real-time and empirical way in the context of achieving the competencies stated in the Turkish Qualifications Framework (TQF). The RLBSD is a holistic and inclusive attempt that would directly contribute to students' language development as the most important life skill, going beyond the limitations of the curricula implemented by MoNE.

Researchers and educators across the country will be able to use the RLBSD in their educational environment with its dissemination/distribution through information technologies. It will be easily accessible through the web, tablet, smartphone, and smartboard applications, and thus take its place in education with an active function. Considering the study's objectives, the RLBSD may serve as a sustainable model as it provides lexicographers a flexible and updatable information platform. The development, implementation, evaluation, and cyclical improvement of the RLBSD constitute this study's main phases.

2.   Methodology

In this section, the methods and techniques, data collection tools, and their analyses are elaborated concerning the literature in line with the study's aim and objective.

2.1   Focus group interview and needs analysis

Data collection techniques used in qualitative research include observation, interview, focus group interview, and document analysis. In this study, a focus group interview (FGI) was employed to determine needs in the development of the RLBSD. FGI aims to obtain in-depth, detailed, and multidimensional qualitative data on the perspectives, experiences, interests, tendencies, thoughts, perceptions, feelings, attitudes, and habits of participants about a set topic (Bowling 2002; Gibbs 1997; Kitzinger 1994, 1995; Krueger 1994; Stewart and Shamdasani 1990).

In this sense, the first phase of developing the RLBSD was to determine users' needs. Focus group interviews were conducted with students, teachers, and lexicographers as the stakeholders of the dictionary to be developed, and details for the lexicographical structure of the RLBSD were determined (Çokluk 2011).

The questions that are planned to be asked in the focus group interview are presented in Figures 1 and 2.

Figure 1:   Questions for the focus group interview

Figure 2:   RLBSD evaluation questionnaire

2.2   Evaluation of the RLBSD by stakeholders

Another data collection used in line with the aim and objectives of the study was a questionnaire. A questionnaire is a technique and/or means for written semi-structured interviews to gather data. It is widely used in social sciences, especially for survey purposes, due to its relative ease of preparation and the possibility of reaching a large number of individuals in a short time. Questionnaires are used to identify/describe a current situation (Erkuş 2013: 161, Büyüköztürk et al. 2016: 124). The questionnaire developed for this study will be used to receive the opinions of lexicographers, teachers, and students on the content and usage of the RLBSD, and the feedback will improve the RLBSD (see Figure 2).

2.3   Evaluating the effectiveness of the RLBSD

One of the current study's objectives is to evaluate the effectiveness of a school dictionary that will be developed based on the methods and principles of corpus linguistics and lexicography. To this end, experimental and control groups were formed from the existing groups/classes by using a reading comprehension test to compare the two groups regarding the effects of the RLBSD. In this sense, the implementation is quasi-experimental. In the quasi-experimental method, existing classes/groups are used in cases where it is difficult or not possible to form experimental and control groups randomly (Robson 1998), so it is also known as the *non-equivalent control group method* (Karasar 2009). In quasi-experimental studies, a control group that is not influenced by the independent variable is used along with the experimental group subjected to this variable (Christensen 2004).

Designed by the researchers involved in the study, the RLBSD will be employed in the experimental group, whereas in the control group, a printed dictionary will be made available to students. The experimental and control groups will include middle school fifth and eighth-graders. In other words, a control group and an experimental group from fifth and eighth grade will be compared within each grade. The reason for choosing the fifth and eighth grades in the study is that they contain students who have just started middle school and those who are about to graduate, and thus, the effectiveness of the RLBSD will be evaluated at different levels. Firstly, peer groups (experimental and control groups) will be formed among the students who have the same level of reading comprehension by administering the comprehension skill test to fifth and eighth-graders, and then it will be revealed whether there is a statistically significant difference in-between after having the experimental group use the RLBSD and the control group a printed dictionary (see Table 1).

Table 1:   Research design

FORMING THE GROUPS	GROUPS	IMPLEMENTATION	EVALUATION
Reading Comprehension Achievement Test	Control 1 (fifth graders)	Text Analysis with Printed Dictionary	Evaluation of Effectiveness
	Experimental 1 (fifth graders)	Text Analysis with the RLBSD	
Reading Comprehension Achievement Test	Control 2 (eighth graders)	Text Analysis with Printed Dictionary	Evaluation of Effectiveness
	Experimental 2 (eighth graders)	Text Analysis with the RLBSD	

2.4   Principles and methods of corpus linguistics and lexicography in the development of the RLBSD

This phase of the study was designed in descriptive and relational models. Descriptive research helps determine a phenomenon as it is, whereas relational research refers to studies in which cause-effect relationships cannot be established, and change and supervision may not be possible either by nature or due to practical reasons (Büyüköztürk et al. 2016; Erkuş 2009; Karasar 2009). Data collection techniques, methods, techniques, and instruments developed for specific purposes are used in research studies (Erkuş 2009). In this project, the methods and techniques introduced by corpus linguistics that can yield empirical results were employed.

The principles and methods of corpus linguistics and lexicography, on which the current study is mainly based, are outlined as follows:

> Today, large-scale corpora that have been developed with their unique features and applications based on written and spoken data from natural language environments make major contributions to language learning and teaching as they contain real-time language data isolated from intuitive examples and secondary data sources (Özkan 2010).

A *corpus* (plural: corpora) can be defined as a body of texts organized based on set standards by selecting written and spoken texts to sample a language following linguistic criteria (McEnery et al. 2006; for other definitions, see Say 2003). The primary criteria for a corpus are its ability to *sample* (sampling) the language in which it is created and its *power* to represent that language. Besides, it should have *a certain limit* (finite size) or not (dynamic size), and be a standard reference for describing the target language in a machine-readable form (McEnery and Wilson 2001).

In the literature, existing corpora vary depending on their aims and foci. In other words, corpora are developed depending on research questions (For

types of corpora containing written and/or spoken texts and that can reflect linguistic diversity in a language, see McEnery et al. 2006).

A set of criteria is considered in the selection of texts to be included in a corpus. The most common ones are taking *literary/academic merit* as the basis, *random selection*, *currency*, *availability*, *demographic sampling*, being *empirical,* and selection from *a broad range of sources*, some of which have advantages and disadvantages over the others (Summers 1993). One or more of these criteria can be used when creating a corpus.

The most crucial issue that would be considered while building a corpus is corpus design. For example, the type of texts included, the number of texts, the selection of texts, sample selection among the pieces of texts, and the sample's length are the measure of whether the design is made consciously (Sampson and McCarthy 2005). Depending on the research question, it is necessary to use a *representative* corpus compiled from the characteristic and typical usage environments of the language to reveal relationalities in language studies fully. Shaping a corpus's quality and content is one of the main issues that make language studies *corpus-based*.

The *power of representation* is a significant defining characteristic of a corpus. This characteristic distinguishes a *corpus* from a *collection of texts* (that are randomly selected). This is a vital practice in corpus design. On the other hand, it is not possible to examine all of a language as a whole. For this reason, forming a sample seems inevitable, and one should be sure of the *representability* of the sample formed for that language while working on it. Furthermore, the balance (the distribution of text types), the quality of the text strata selected for each type, and the sample are among the main elements that constitute the *representative power* of a corpus (McEnery et al. 2006).

Today, *representativeness* that involves *text-type distribution/distributiveness* and *sample selection/sampling* is a prerequisite for corpus applications. On the other hand, *representativeness* comprises a sample with all the variables in a population. In studies on sample size, researchers agree that representativeness, which refers to how many texts a corpus should contain and how many words a text should contain, is the most important decision-making point. Accordingly, the variety of texts in a language and their linguistic distributions are among the issues that should be considered in corpus design (Sampson and McCarthy 2005).

Considering the arguments mentioned above and corpus characteristics, the corpus used for the Real-life-based School Dictionary can be regarded as *a specialized corpus*. The specialized corpus to be created contains two main subcorpora. The first sub-corpus is the *Corpus of Turkish Children's Literature*, a specialized corpus that focuses on the literary works for Turkish children in the context of qualitative and quantitative research topics. *This corpus was created within the scope of another national project funded by TÜBİTAK.* With the project in question, Turkish Children's literature was used as the research population for various analyses such as *the readability and age-appropriateness of children's books,*

lexical variety and lexical domain patterns, morphological, lexical, and syntactic features, and internal and external structural features. This specialized corpus of "Turkish Children's Literature" was compiled according to corpus linguistics principles and methods and contains 8,639,522 million words. This corpus containing 1,089 texts in different types including texts from children's periodicals, will be evaluated in the proposed Real-life-based School Dictionary. As for the second sub-corpus, it includes MoNE-approved coursebooks. The research population was completed by digitizing the coursebooks based on corpus design steps.

2.4.1 Morphological, lexical, and syntactic structure of the corpus to be developed for the RLBSD

The texts that included in the corpus were sententially parsed, and metadata for every sentence such as sub-corpora and text types [*Texts of Turkish Children's literature* (novel, poem, story, memoir, periodicals, etc.), *MoNE-approved coursebooks* (by grades and courses), *Other Types* (other texts that children may encounter)], author, work title and year of publication will be tagged and added to the corpus.

At this stage, lemmatization was performed on the corpus, and the lemmas were deduplicated.

2.4.1.1 An example of a sentence and sentence tagging

< Tales was a great scientist who calculated the height of the Egyptian pyramids and "drew a right-angled triangle into a circle". >
                                                 <Corpus of Turkish Children's Literature> **First sub-corpus**
                                        <Essay> **text type**
                                            <mavisel yener> **author name**
                                                  <Pyramids> **work title**
                                                      <2009> **year of publication**

2.4.1.2 An example of morphological analysis[1]

Mısır [???] | *pramitlerinin* [piramit(pyramid)+Noun+A3pl+P3sg+Gen] |
yüksekliklerini [yüksek(height)+Adj^DB+Noun+Ness+A3pl+P3sg+Acc] |
hesaplayan [hesapla(calculate)+Verb+Pos^DB+Adj+PresPart] | *ve* [ve(and)+Conj] *bir*
[bir(a)+Det] | *dairenin* [daire(circle)+Noun+A3sg+P2sg+Gen] | *içine*
[iç(in/into)+Noun+A3sg+P3sg+Dat] | *dik* [dik(right-angled)+Adj] | *üçgen*
[üçgen(triangle)+Noun] | *çizen* [çiz(draw)+Verb+Pos^DB+Adj+PresPart] | *büyük*
[büyük(great)+Adj] | *bir* [bir(a)+Det] *bilim* [bilim(science)+Noun] | *adamıydı*
[adam(man)+Noun+A3sg+P3sg+Nom^DB+Verb+Zero+Past+A3sg] | *Tales* [???]

2.4.3   Data processing in the RLBSD

Data processing was performed via an online platform. The data processing platform created is a previously experimented platform by which other studies based on the Turkish language have also been conducted. Various lexicography applications supported by TÜBİTAK and university funds were carried out on the said platform (http://turkcederlem.mersin.edu.tr/) (see Figure 3).
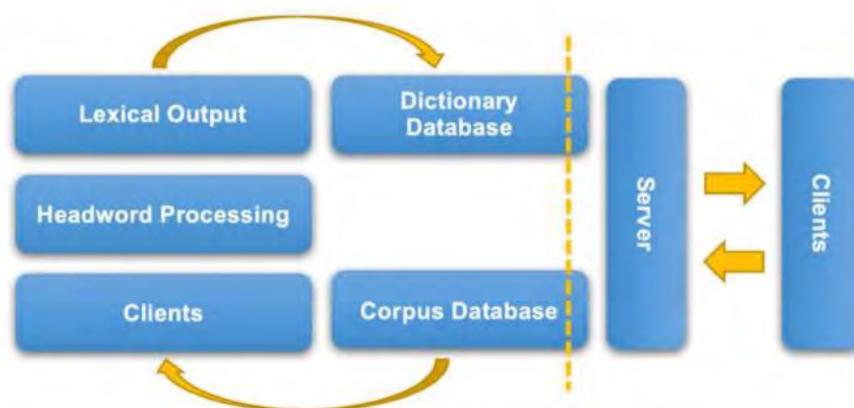
Figure 3:   Data processing platform (Özkan 2013)

In the 'data-processing platform' shown above, two separate databases on the server provide an uninterrupted service for users to process and view data. Simply, users make lexical queries on the corpus database and then perform data processing on the query results, followed by saving entries on the dictionary database. The saved and processed data are published instantly on the server with query screens.

2.4.3.1 Data processing in the RLBSD over the corpus

The RLBSD is subjected to the processing of tagging and reporting over a system represented in Figure 4.
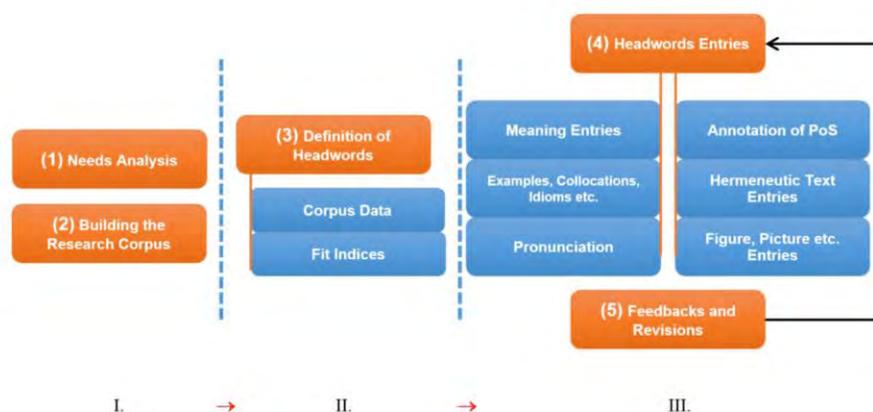
Figure 4:   Data processing stages

In brief, the proposed project includes five main steps in three phases. *The first phase involves* (1) determining the needs and (2) creating the corpus, *the second phase* (3) determining the lexical entries, and *the third phase* (4) structuring the lexical entries which include meaning entries, word type tags, selection of examples, identifying collocates and related idioms, pronunciation, reader-friendly text, figures and visuals, and (5) feedback and corrections.

These steps along with the methods and techniques to be implemented in the projects are explained in the following:

(1)   Determining the needs

In brief, determining the needs of the target population is one of the most decisive elements when developing a dictionary. From this perspective, the needs of the target population and/or stakeholders must be determined precisely. Focus group interviews with subject-matter teachers and linguistics/lexicography experts will be employed in this step (see the methodology section above).

(2)   Creating the corpus

The corpus that would be used for the Real-life-based School Dictionary is a specialized corpus. The specialized corpus created contains two main sub-corpora. The first sub-corpus is the specialized Corpus of Turkish Children's Literature (see http://turkcederlem.mersin.edu.tr/cocuk/ and Özkan 2014). As for the second sub-corpus, it includes MoNE-approved coursebooks and the periodicals that children may encounter.

(3)   Determining the lexical entries

When the history of lexicography is examined, it can be observed that the language samples that linguists and lexicographers used in their attempts are mostly based on native-speaker intuitions (i.e., armchair lexicography), and these works have yielded results that are not right, and/or that could not fully reflect linguistic realities. Moreover, individual preferences and intuitive approaches of linguists and lexicographers necessarily mean differing results for every language expert (McEnery et al. 2006). In current lexicographical traditions, practices such as collecting examples, using individual knowledge, and forming index cards are methods that are accepted as non-functional in developing a dictionary (Atkins and Levin 1995). On the other hand, school dictionaries are important resources in first language education, and studies on how and from which sources these dictionaries should be developed have been widely addressed in the literature (Bergenholtz and Gouws 2012; Malkiel 1967).

In a lexicographical study, all steps such as *defining linguistic components, determining the frequencies of occurrence for these components, forming concordances,* and *extracting collocational structures* provide quality and valid data for the researchers in the selection of dictionary entries. In this respect, dictionary entries in the RLBSD are structured in the light of scientific and experimental methods.

While determining lexical entries, there are two ways that are followed. The first of these is to obtain lists of lexical entries from the lexemes that occur in the corpus. This is one of the standard data processing steps on the corpus platform. Regarding the vocabulary retrieved from the MoNE-approved coursebooks that can be lexical entries in the RLBSD, structures are determined as lexical items employing simple fit index calculations after asking for the opinions of three subject matter teachers for each coursebook, in addition to frequency and distribution criteria. Furthermore, since the corpus will be lemmatized the words that would be retrieved through the *simple fit index* calculations and those obtained from the corpus will be structured as lexical entries.

Thus, the lexical entries will be those: **a.** that are obtained from the corpus, and **b.** that are retrieved through simple fit index calculations based on the opinions of field experts. This type of process means that the selection of lexical entries will be done with an empirical approach.

While identifying the subject areas (i.e., science, social studies, or mathematics in the context of courses), the metadata about the texts included in the corpus such as *sub-corpus, text type, author name, work title,* and *date of publication* can be automatically obtained through the corpus formed digitally. This and similar metadata will form the basis for further research as standard corpus output that can provide researchers the data that can be used in the detailed analyses of the corpus.

(4)   Structuring the lexical entries

When creating a dictionary, a set of lexical entry configurations is used in accordance with its purpose. These are *writing* (spelling), *pronunciation*, *inflections*, *word forms* (part of speech), *meaning(s)*, *definitions*, *examples*, *usage*, *other derivations* (run-ons), and *root information* (etymology) (Jackson 2002; Hanks 2003). Other configurations (picture, sound file, etc.) can be added to these by using information technologies. In the structuring phase, what a dictionary entry includes is decided based on the lexicographer's aim. The data obtained from corpora form the basis for this decision.

In other respects, as mentioned before, although school dictionaries have a similar structural feature with general dictionaries (entry, meaning, part of speech, example, etc.), it has to contain certain configurations specific to the target audience, unlike general dictionaries. The basic arrangements to be included in the RLBSD and their sources can be summarized as follows:

a.   Meaning/definition arrangement

Experts of linguistics/lexicography arranged the meanings to be provided for lexical entries in a context-sensitive way. At this stage, the target population was considered, and the meaning arrangements were shaped by making use of the meaning patterns existing in general dictionaries. The corpus's most significant contribution to this study is that all lexical contexts can be seen, inspected, and included while assigning meanings to lexical entries. This is of significance in making a simple, clear presentation of meaning in line with the target audience's needs by making comparisons in connotations. The development of specialized dictionaries with a corpus-based or corpus-driven approach is also important for words to be presented in a semantic harmony (Yu and Cai 2009).

As stated in the literature (Günay 2007), the existing definitions in lexicography (e.g. *definitions about meaning and extension, descriptive meaning, operational definition, explicative definition,* and *conventional definition*) are formatted depending on the characteristics of lexical entries in the stage of meaning and definition. When providing definition and meaning, for example the lexeme "**ak**" can be defined as "beyaz" (white) in general dictionaries, whereas its definition can be "Kar veya süt renginde olan." (the color of snow or milk) in a school dictionary. Such an approach can present a meaning/definition directly to users, considering the dictionary's target audience.

b.   Pronunciation

Another arrangement to be included in the RLBSD in accordance with its purpose is the pronunciation of the lexical entries. Audio recordings were added to the entries of the RLBSD.

c.   Determining the part of speech and conceptual field

The part of speech was tagged by sampling all uses of words in a context-sensitive way. This is a step that can be implemented easily through corpus linguistics applications.

d.   Selection of examples

In corpus queries, all contextual patterns about the lexical entry being searched can be viewed and listed for selection. In this sense, linguists/lexicographers can choose the right and functional examples in accordance with the purpose of the dictionary and have the opportunity to configure the entries.

e.   Determining the collocational and idiomatic expressions

Revealing collocational and idiomatic expressions out of a corpus by using n-grams means finding out the main structures in the formation of meaning patterns. In corpora, the retrieval of multi-word combinations [... n-3, n-2, n-1 || n+1, n+2, n+3 ... (bigram, trigram...)] is one of the standard processing steps, the tagging of the listed constructions simplifies the work of linguists/lexicographers as they identify collocational and idiomatic expressions.

f.   Explanatory texts

Another characteristic of the RLBSD is enriching the entries with supportive/explanatory texts that explain them. At the end of some entries, additional information about the headword can be found. This information includes extra information about a place, a person, a thing, a process. With this explanatory information, the meaning of the headword becomes clearer for the dictionary users.

g.   Figures, images, etc.

In accordance with the purpose and the target audience, visuals containing, explaining, and representing the lexical entries were used. Accordingly, the list of lexical entries determined by lexicography experts through simple fit indices was visualized with images and figures.

h.   Vocabulary teaching, vocabulary games, tests for teaching spelling

Particularly concerning vocabulary teaching, certain digital applications (e.g. matching, comparing, completing), that are widely used in language education will take their place in the corpus-based entry configurations.

Consequently, a sample lexical entry can be represented as shown in Table 2.

Table 2:   Sample of lexical query of "üçgen"
(see also http://okulsozlugum.com/arama/ and
http://turkcederlem.mersin.edu.tr/okulsozluk/)

(5)   Feedback and corrections

Apart from corpora, another way an individual develops a general dictionary that can be used as a data source is through user suggestions. User suggestions may include the addition of missing words or word meanings to the dictionary etc. Although dictionary developers may have a vast and comprehensive database, they will always be outnumbered by those who use the dictionary. This means that it is likely for a dictionary user to not find a word they are looking up in a dictionary. No matter how many and diverse the texts that dictionary makers based their dictionary on, users always come across more texts. In this regard, the developer of a general dictionary who is open to user contributions and suggestions should be able to benefit from user feedbacks (Bozkurt 2017) meaning that user practices and preferences can also be taken into consideration to develop a more useful dictionary.

This project has an interactive and multi-stakeholder construct. Merely having the opinions of lexicographers would not be sufficient for developing a school dictionary. User practices and preferences influence the stages of dictionary development with user feedback and corrections. Necessary arrange-

ments and corrections will be made by receiving feedback from subject matter teachers, lexicographers, and students.

It is also common practice to log user queries [access logs] in modern corpus-based and IT-supported dictionary applications. Today, advanced and statistical software tools can provide the researchers data on which lexemes users search for as well as their user profiles (e.g. age, region, gender, etc.). The most well-known of these are open-access applications such as Google Analytics and Yandex Analytics. In this respect, relevant information technologies can be used to receive feedback and make corrections.

3.   Conclusion

In the current project, RLBSD differs from the school dictionaries that have been introduced so far in terms of both its creation method and the possibilities of reaching the end-user and/or accessing the dictionary. First of all, the focus group interview for dictionary creation is a new practice in itself, at least for Turkish lexicography. Besides, the use of "a special field corpus" that includes vocabulary elements that school-age children encounter can be considered the most appropriate approach to the content organization of a school dictionary. In this regard, it can be concluded that the use of corpus linguistics principles and methods and lexicology principles together in the creation of RLBSD can be considered a breakthrough in the Turkish lexicography tradition. Although creating a dictionary may seem like a lexicographer's job, a dictionary cannot be considered independent of user evaluation. From this perspective, the stages of evaluating the dictionary's effectiveness after the creation of RLBSD in educational environments are considered a brand-new approach that corresponds to an application that has not been experienced before in Turkish lexicography.

As is known, the corpora provide important contributions to the lexicology field with the experimental results they provide. When considered in this regard, corpus linguistics applications can be accepted as an empirical criterion for dictionary creation. From this perspective, RLBSD draws attention as a source that contains real-time data in terms of item selection and item configurations. With this feature, RLBSD can be placed on an axis outside the traditional lexicography understanding.

It can be said that the use of and access to RLBSD developed with the aforementioned experimental methods may become a standard for corpus-based dictionary creation due to the use of information technologies used in its e-spread and distribution. In this sense, RLBSD with its feature of ease of access through the Internet, tablet, smartphone, and smartboard applications will make it an effective resource as course material in educational environments. Also, RLBSD can be considered a sustainable model as it offers a flexible and updatable platform. In addition, since the effectiveness of RLBSD is to be evaluated, how it affects learning outcomes will also be tested.

Endnote

1.   Every lexical item in the corpus to be created will be analysed using the morphological analy-
     sis tool (see http://www.hlst.sabanciuniv.edu.tr/TL/) developed by Kemal Oflazer.

References

Akcan İ.P. 2012. Türkçe Öğretiminde Sözlük Kullanımı: Türkçe Çalışma Kitaplarında Sözlüğe
     Gönderen Alıştırmalar. Aksan, M. and Y. Aksan (Eds.). 2012. *Türkçe Öğretiminde Güncel
     Çalışmalar*: 93-102. Ankar: Şafak Yayınları.
Atkins, B.T.S. and B. Levin. 1995. Building on a Corpus: A Linguistic and Lexicographical Look a t
     Some Near-synonyms. *International Journal of Lexicography* 8(2): 85-114.
Atkins, B.TS. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography.* Oxford/New York:
     Oxford University Press.
Bergenholtz, H. and R.H. Gouws. 2012. What is Lexicography? *Lexikos* 22: 31-42.
Bowern, C. 2008. *Linguistic Fieldwork: A Practical Guide.* New York: Palgrave Macmillian.
Bowling, A. 2002. *Research Methods in Health: Investigating Health and Health Services*. 2nd edition.
     Buckingham/Philadelphia: Open University Press.
Bozkurt, F. 2017. *Sözlükselleşme: Genel Sözlükler için Sözlük Birim Seçimi.* Istanbul: Kesit Yayınları.
Büyüköztürk, Ş. et al. 2016. *Bilimsel Araştırma Yöntemleri.* 22nd edition. Ankara: PegemA Yayıncılık.
Christensen, L.B. 2004. *Experimental Methodology.* 9th edition. Boston: Pearson Education.
Çokluk, Ö. 2011. Nitel Bir Görüşme Yöntemi: Odak Grup Görüşmesi. *Kuramsal Eğitimbilim* 4(1): 95-107.
Çotuksöken, Y. 1999. Okul Sözlükleri Üzerine. *Kebikeç* 7–8: 61-65.
Erkuş, A. 2009. *Davranış Bilimleri için Bilimsel Araştırma Süreci.* Ankara: Seçkin Yayıncılık.
Erkuş, A. 2013. *Davranış Bilimleri için Bilimsel Araştırma Süreci.* 4th edition. Ankara: Seçkin Yayıncılık.
Gibbs, A. 1997. Focus Groups. *Social Research Update* 19(8): 1-8.
Günay, D. 2007. *Sözcükbilime Giriş.* Istanbul: Multilingual.
Hanks, P. 2003. Lexicography [Chapter 3]. Mitkov, Ruslan (Ed.). 2003. *The Oxford Handbook of Com -
     putational Linguistics*: 48-69. Oxford: Oxford University Press.
Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography.* London/New York: Routledge.
http://turkcederlem.mersin.edu.tr/. *Date of access:* 17-12-2019
http://turkcederlem.mersin.edu.tr/cocuk/. *Date of access:* 17-12-2019
http://turkcederlem.mersin.edu.tr/okulsozluk/. *Date of access:* 5-10-2020
http://okulsozlugum.com/arama/. *Date of access:* 16-5-2021
http://www.hlst.sabanciuniv.edu.tr/TL/. *Date of access:* 9-02-2009
Jackson, H. 2002. *Lexicography: An Introduction.* London/New York: Routledge.
Karasar, N. 2009. *Bilimsel Araştırma Yöntemi.* 20th edition. Ankara: Nobel Yayın-Dağıtım.
Kitzinger, J. 1994. The Methodology of Focus Groups: The Importance of Interaction Between
     Research Participants. *Sociology of Health and Illness* 16(1): 103-121.
Kitzinger, J. 1995. Qualitative Research: Introducing Focus Groups. *British Medical Journal* 311: 299-302.
Krueger, R.A. 1994. *Focus Groups: A Practical Guide for Applied Research.* 2nd edition. Thousand
     Oaks, CA: Sage.

Malkiel, Y. 1967. A Typological Classification of Dictionaries on the Basis of Distinctive Features. Householder, Fred W. and Sol Saporta (Eds.). 1967. *Problems in Lexicography*: 3-24. Bloomington: Indiana University.

McEnery, T. and A. Wilson. 2001. *Corpus Linguistics: An Introduction.* 2nd edition. Edinburgh: Edinburgh University Press.

McEnery, T. et al. 2006. *Corpus-Based Language Studies: An Advanced Resource Book.* London: Routledge.

Melanlıoğlu, D. 2013. Ortaokul Öğrencilerinin Sözlük Kullanma Alışkanlıkları: Nitel Bir Araştırma. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi* 2(2): 266-284.

Özkan, B. 2010. Türkçenin Öğretiminde Sıfatların Eşdizim Sözlüğü: Yöntem ve Uygulama. *e-Uluslararası Eğitim Araştırmaları Dergisi* 1(2): 51-65.

Özkan, B. 2013. Yöntem ve Uygulama Açısından 'Türkiye Türkçesi Söz Varlığının Derlem Tabanlı Sözlüğü'. *bilig. Türk Dünyası Sosyal Bilimler Dergisi* 66(3): 149-178.

Özkan, B. 2014. Bir Özel Alan Derlemi Olarak 'Türk Çocuk Yazını Derlemi' ve Türk Çocuk Yazını Üzerine İleri Uygulamalar Projesi. Özkan, B. (Ed.). 2014. *Çocuk Yazını Araştırmaları-I:* 1-40. Adana: Karahan Kitabevi.

Özkan, B. et al. 2019. The Method of the Real Life Based School Dictionary. Gürlek, M. et al. 2019. *ASIALEX 2019. Proceedings of the 13th International Conference of the Asian Association for Lexicography, Istanbul*: 94-105. Istanbul: Asos.

Robson, C. 1998. *Real World Research: A Resource for Social Scientists and Practitioners.* Blackwell.

Sampson, Geoffrey and Diana McCarthy. 2005. *Corpus Linguistics: Readings in a Widening Discipline.* London/New York: Continuum.

Say, B. 2003. *Türkçe için Biçimbirimsel ve Sözdizimsel Olarak İşaretlenmiş Ağaç Yapılı Bir Derlem Oluşturma.* Proje No: 199E026. Proje Raporu-Ek-b: Türkçe Ağaç Yapılı Derlem Kılavuzu. Ankara.

Stewart, D.W. and P.N. Shamdasani. 1990. *Focus Groups: Theory and Practice.* Newbury Park, CA: SAGE.

Summers, D. 1993. Longman/Lancaster English Language Corpus — Criteria and Design. *International Journal of Lexicography* 6(3), 181-208.

Yu, P.-F. and J.-C. Cai. 2009. Semantic Prosody: A New Perspective on Lexicography. *US-China Foreign Language* 7(1): 20-25.

Fostering Learners' Online Dictionary Skills through Active Dictionary Rubrics

Mari Carmen Campoy-Cubillo, *Department of English Studies, Universitat Jaume I, Castelló de la Plana, Spain (campoy@uji.es)*

Abstract: This article aims to explore new ways to improve dictionary skills for language learners by promoting the use of *active dictionary rubrics*. It also tries to get a deeper understanding on how students perceive dictionaries before and after this type of training in dictionary use. To this end, a total of 75 students participated in the study and answered a questionnaire on dictionary use previous to the experiment. A second questionnaire was delivered after the instruction period to compare the users' perception of dictionary affordances. A mixed-method approach is used to collect quantitative and qualitative data via both questionnaires. This study reports on students' dictionary content awareness and how this awareness may increase by creating a habit in dictionary use routines.

Keywords: DICTIONARY USE, DICTIONARY USERS, DICTIONARY RUBRICS, ENGLISH LANGUAGE TEACHING (ELT), DICTIONARY SKILLS, ONLINE DICTIONARIES, DICTIONARY CONSULTATION, WRITING SKILLS, LEARNERS' DICTIONARIES, DICTIONARY USE TRAINING

Résumé: Favoriser les compétences des apprenants pour l'utilisation de dictionnaires en ligne à travers les rubriques actives. Cette étude vise à explorer de nouvelles façons d'activer les compétences dans l'usage des dictionnaires pour les apprenants de langues, en favorisant l'utilisation de ce que nous avons appelé des en-têtes actives d'utilisation de dictionnaire. Elle essaie également de mieux comprendre comment les élèves perçoivent les dictionnaires avant et après ce type de formation. À cette fin, un total de 75 étudiants ont participé à l'étude et ont répondu à un questionnaire sur leur utilisation du dictionnaire avant l'expérience. Une approche mixte est suivie dans laquelle certaines questions sont numériques tandis que d'autres sont qualitatives. Un deuxième questionnaire est livré après la période d'instruction pour comparer la perception des utilisateurs sur les possibilités des dictionnaires. Cette étude rend compte de la sensibilisation au contenu du dictionnaire et comment ces connaissances peuvent augmenter si on crée une habitude dans les routines d'utilisation du dictionnaire.

Mots-clés: UTILISATION DE DICTIONNAIRES, UTILISATEURS DE DICTIONNAIRES, EN-TÊTE, ENSEIGNEMENT DE LA LANGUE ANGLAISE (ELT), COMPÉTENCES D'UTILISATION DES DICTIONNAIRES, DICTIONNAIRES EN LIGNE, CONSULTATIONS DE DICTIONNAIRES, COMPÉTENCES D'ÉCRITURE, DICTIONNAIRES POUR LES APPRENTIS, FORMATION POUR L'USAGE DU DICTIONNAIRE

0.   Introduction

Dictionaries play an important role as an aid in the process of learning a language. In this sense, ability to use a dictionary is generally believed to yield better results in the language learning process as well as to foster learners' autonomy. In the area of dictionary skills, research focuses mostly on dictionary use with different purposes and in different contexts (Rundell 1999, Schofield 1999, Laufer 2011, Müller-Spitzer 2014, Liang and Xu 2018), the effectiveness of different dictionary designs, the role of outer texts in dictionary pedagogy, and information presentation options (Nkomo 2015, Dziemianko 2016, 2019, Ptasznik and Lew 2019, Ptasznik 2018). But the majority of studies on dictionary use do not tackle how dictionary users are trained to become proficient users or how the training may affect test results. That is, dictionary use is analysed without considering previous specific dictionary training methods of users or the lack of them. These analyses pay attention to how the use of dictionaries in a particular task may improve results as compared to not using dictionaries. This could be compared to testing a group of students on writing skills without considering their initial writing competency level. If the group has students with different language proficiency levels, it is logical that those with a high writing skills level will perform better than those with a lower level.

1.   Literature review

This section deals with research related to dictionary use and dictionary skill training. Some studies on dictionary use directly address dictionary skill training (Bishop 2001, Nation 2001, Bogaards 2003, Carduner 2003, Hadebe 2004, Chi 2003, Lew and Galas 2008, Lopera 2019). Hadebe (2004) focuses on dictionary skill training in relation to the dictionary culture of the Nembele community in Zimbabwe. He identifies teacher training in dictionary skills as the first condition so that learners may receive the most adequate training according to their specific needs. Bogaards (2003: 28) points out the necessity for dictionary training and how its lack of development is in sharp contrast with the progress observed in dictionary quality. He states that:

> One of the recurrent themes in all of these publications is that dictionaries have been improving considerably over the past fifteen years but that instruction in dictionary use remains essential if users want to take advantage of the real riches of their dictionaries. (…) It is remarkable that in most teacher training programmes no time is set aside for dealing with dictionary use, just as in most language programmes in schools no attention is paid to dictionary instruction.

Nation (2001) is more precise on how this should be done, and distinguishes six important strategies that should be considered for dictionary training: evaluating the need to look up a word, finding the entry, reading the entry, relating the meaning into the context, changing the word into its basic form and guess-

ing the meaning from context. Lopera (2019) uses these strategies to conduct dictionary use training in reading comprehension lessons and supports them with class and homework tasks as well as with field notes. Lopera states that the combination of reading strategies with dictionary training yielded positive results regarding metacognitive awareness, selective use of dictionary and improvement in the students' reading skills.

Chi (1998) designs a full course of academic writing in which for six weeks specific contents of dictionary use training are included (Chi 1998: 577). A more detailed program is illustrated in Liu (2014) explaining the contents of nine training lectures on the following topics, namely:

1. How to choose and use English dictionaries.

2. Introduction to some commonly used English dictionaries.

3. Using monolingual dictionaries in English study.

4. Using dictionaries to gain improvement in reading and vocabulary.

5. Using dictionaries in translation from English into Chinese.

6. Using dictionaries in translation from Chinese into English.

7. Writing conference (assisted by dictionaries).

8. Using semantic information in dictionaries.

9. English writing with the help of dictionaries

Procedures followed for each of the nine lectures are given (Liu 2014: 2139-2141) and each session was also used as feedback for the following lectures. After the training the researcher conducted a half-structured interview. Interview results detect that dictionary training had a positive influence in that the learners were able to observe word usage and collocational behaviour and that they were better at distinguishing synonyms. This researcher also underlines the role of the teacher in the whole training method and the necessity to train the teacher into dictionary skill teaching.

Bishop (2001), Carduner (2003), Lew and Galas (2008) and Ali (2017) provide pedagogical proposals after designing and implementing dictionary training courses. Bishop (2001) designed a self-study dictionary training course to assess the value of such training for writing skills. Accuracy and language quality were tested comparing intra and inter-student writing skills competence. Results showed improvement in both areas and pedagogical proposals direct towards the use of intensive correction and comparison procedures. Based on the results of a dictionary training course for Spanish students, Carduner (2003) discusses which dictionary skills should be included in dictionary skills courses and provides sample dictionary exercises used in the study. Lew and Galas (2008) place the emphasis on presenting and practicing different dictionary skills in 12 dictionary skills sessions with Polish final year primary school children. Their

study reports on the findings regarding students' improvement for each specific area of dictionary skills tested and includes samples of material and tests used for the study.

Following Tono (1989) and Yorkey (1982), Ali (2017) designs a 15-module dictionary skills course for Sudanese English majors to study the ability of students in their use of dictionaries before and after the training. The dictionary skills test used in the research is provided in this article. This is another important aspect that should not be overlooked in the field of dictionary skills training: the availability of tests and materials used as part of the research. If such material is given, a more detailed consistent research may be carried out by researchers working in this field of knowledge. Obviously, the possibility to replicate studies on dictionary skills training depends on how much a specific method and materials are made public and available.

Ranalli (2013a, b) differs from the previously mentioned studies in that he develops an online training program for learning vocabulary and developing dictionary skills. As such, this researcher bases his training in the multimodal learning theory and considers instruction from a multimodal perspective to which cognitive load theories, and different types of knowledge representation are applied. Tutorials, videos and multimodal exercises are integrated in this online training program and at different points there are links that take the trainee to a selected online dictionary. The efficiency of the prototype course (Ranalli 2013a) is tested with tertiary education students (Ranalli 2013b) and reports, among other issues, the improved ability of participants in selecting the best dictionary for a particular purpose and ability to correct the types of errors for which they had been trained in the course by using dictionaries.

As shown above, some research is carried out in an effort to include dictionary training and use in the curriculum. However, the majority of national curricula do not itemise the aspects of dictionary training that should be developed in the classrooms. The Common European Framework of Reference mentions dictionaries and reference materials but does not analyse their role in the language learning process. An exception in this panorama is the case of dictionaries in the South-African curricula (Nkomo 2015), which clearly outline dictionary use and training in relation to educational level and subject where they should be introduced, type of dictionary to be used and purpose of use, and the activities that may be developed in each case. Two recent studies need to be included in relation to dictionary skills: Egido Vicente (2021) who thoroughly discusses the value of dictionary as a language learning resource, lexicographical skills and their relation to curricula design, and Chi (2020: 79) who clearly evaluates the relationship between lexicography and language methodologies:

> As EFL lexicography and teaching methodologies have been evolving on their own paths in the past decades, the relationship between the two areas has become less dynamic and supportive (complementary). The broken relationship has gravely impacted the use of the dictionary among EFL learners who follow a

structured language curriculum. One suggestion to bridge the gap between the dictionary and EFL teachers is to provide training to English language teachers on "dictionary literacy".

This study presents a new proposal for dictionary skills training that is embedded in the English language subject in the first year of an English for Special Purposes (ESP) university course. The experiment was developed during an academic semester in which 75 students participated and received instructions to better use online dictionaries. To this end, three online dictionaries were used as part of the training: *Cambridge Dictionary*, *Macmillan English Dictionary* and *Merriam-Webster Dictionary*. The bilingual sites (English–Spanish/Spanish–English) for the last two dictionaries were also employed. The following section explains the methodology followed and provides illustrative examples of the material employed.

2.   Methodology

In this study, a new type of rubric — called *active dictionary rubric* — was designed by the author. Active dictionary rubrics start out from the concept of holistic rubrics. Holistic rubrics use the same criteria to assess overall achievement of a number of items. The different items in the rubric are assigned specific scores. Thus, Table 1 could be an example of a holistic rubric in which we aimed at assessing dictionary skills (Campoy-Cubillo 2015) in relation to the ability to locate, understand, interpret, evaluate, record, and implement information provided in dictionary definitions (DEF) and examples (EX):

Criteria	Score: ability to implement information in the task at hand
(DEF) Locate the required entry, find relevant word sense, interpret definition and evaluate which part is useful, note down information to apply to a given task.	50%
(EX) Locate examples provided for a word/word sense, understand word use in context provided by the examples, interpret rules of use, evaluate how to apply to own sentence, note down relevant information.	50%

Table 1:   Example of dictionary skills holistic rubric

The holistic rubric evaluates performance in relation to defined criteria. But before these criteria are met, students need specific training. Otherwise, we would be assessing the individual's previous knowledge and skills instead of

the learning process or the effectiveness of a particular dictionary design and affordances without depending on the teacher's knowledge and experience or evaluating the quality of the dictionaries employed.

Thus, in any stage of assessment process, it is important to ensure that the assessment instrument responds to the training methods that prepare the learner for that particular assessment type. The challenge in teaching dictionary skills is that, in contrast to other language learning skills, there are no dictionary training methods that are clearly outlined nor dictionary training materials that may be easily found, and the literature on dictionary instruction is scarce. In order to assess dictionary skills in a fair and effective way, dictionary skills training methods and proposals are necessary. In the field of dictionary instruction, *active dictionary rubrics* are a step in this direction.

Following Nation (2001: 296), who stated that "Dictionary use is a kind of language-focused learning: the deliberate, explicit study of words" *active dictionary rubrics* are designed to make dictionary use "deliberate and explicit". This means that students are not expected to make an unguided one-time dictionary look-up for a word they choose to search. On the contrary, through the use of active dictionary rubrics they are guided to perform planned and repetitive word look-up activities. They also use dictionaries in a way that is clear and direct and in connexion with the writing task they are working on in class.

In this study, students were asked to deliver "micro-writing tasks". These are short texts that they are requested to write, responding to a basic question, for example: "When was the last time you felt anger?" or "When was the last time you felt frightened?" or "Do you remember a time when you really enjoyed yourself?" Students answer this question with a maximum of five sentences. The teacher corrects the writings and studies which mistakes can be corrected with the help of a dictionary. Student sentences are numbered (S1, S2, S3, S4, S5) and the mistakes are underlined in each sentence. The items in the first column of the rubric (see Table 2) state the type of information look-up process that will be helpful to correct the mistake. The first row in the rubric indicates the number of the sentence where the mistake appears. An active dictionary rubric can be created with the sentences of one micro-writing task presented by an individual, or with 4–5 sentences from different micro-writing tasks from several students. These rubrics can be used with all students, whether they wrote these sentences or not, because the aim is to interrogate the dictionaries in order to become aware of how and where to find information on words. The rubrics can also be used in the following academic years with a different group of students. Not all categories need to be used every time an active dictionary rubric is provided. The irrelevant ones may be left blank or can be removed if there is no item in the essay that fits the category. Appendix 1 shows a blank rubric with all the categories used in this study. Designing rubrics to work with other dictionaries may entail creating new categories. The number of sentences may also be changed depending on the amount of time a teacher wants to spend on this type of training.

The rationale behind the suggested dictionary actions are based on teacher knowledge of the students' difficulties as well as teacher's dictionary knowledge. Students' mother tongue also plays an important role. For instance, "I'll take care of your tent is safe" is a literal translation from Spanish, and the teacher knows what the student meant. This is how the teacher can think about the ways the dictionaries employed will provide an answer. Likewise, the student wrote "tent" because it sounds similar to "tienda" (shop). If they read the definition for "tent" they will see that it has nothing to do with the word they were trying to find, and then search again for a better result. The same happens for the use of collocations, teachers may identify wrong collocations and address the student to the place in the dictionary where they will find the answer. It is also frequent that students use dictionary translations without checking whether that translation is adequate in the sentence they are using, or that they select a similar word that is semantically related but does not fit a sentence. They are then guided for instance to read definitions (or compare two definitions of related words) and pay closer attention to word meaning to make informed choices. They may also be told to read example sentences for the same purpose.

Dictionaries used in the training were monolingual English dictionaries: *Macmillan English Dictionary Online, Cambridge Dictionary Online* and *Merriam-Webster Dictionary Online*, and Spanish–English/English–Spanish bilingual options in *Cambridge* and *Merriam-Webster (Spanish Central) Online* Dictionaries.

MICRO-WRITING TASK

Q: *If you were a police officer, what would you say to someone who has been robbed?*

Sample of a student's answer to this micro-writing task:

S1. Everything will return to normal. It is best to stay calm. I assure you, that personally I'll <u>take care</u> of your <u>tent</u> is safe.

S2. We understand your <u>frustration</u>, but in fact there is nothing to <u>care about</u> because we already arrested the robbers.

S3. We would like to inform you that you can rest easy knowing the thieves are on <u>jail</u>.

S4. Let me give you <u>a good advice</u>: it will be better to remain <u>calm</u> all the time and also be positive.

S5. I <u>wish</u> that this issue could be resolved soon so that we can feel <u>more safer</u>.

Micro-writing active dictionary rubric	S 1	S 2	S 3	S 4	S 5
(DEF) Definition: check definition for a better understanding/use of word	Read entries for "take care of". Is that the best choice for this sentence?	Read definitions for "care about". Do they fit sentence 2?			

494   Mari Carmen Campoy-Cubillo

	Check other entries for "take care" followed by different words (that, with … etc). Read the definition for "tent". Is that the word you wanted to use here? If not, which word should you use instead?				
(SENS) Check you have chosen the correct sense of the word				Provide a sentence of your own with 2 of the different senses of the word "calm"	
(COL) Check, learn or use correct collocation		Find the collocations for "frustration" in Macmillan, read the examples and decide which one means "lleno de frustración"i. Find two adjectives collocating with "frustration" expressing degree of frustration.	Which prepositions can you use with the word "jail" (noun)?		
(EX) Check examples for a more natural use of word(s) / expressions (click on "more examples" if you see this link in the dictionary)		Read the examples for "care about" and "worry about". Which verb fits better in sentence 2?			Read the examples for the verb "wish". Rewrite the sentence or find a better choice using the Spanish–English dictionaries.
(GRMM) Read the grammar information about this word					See Word Forms for "safe"

(US) Usage / get it right				Read the information for "advice" in the section "Get it right" inside Macmillan.
(IDM) Look at the "More meanings" tab to find idioms, phrases and phrasal verbs related to this word to learn new expressions	Cambridge: see idioms with "return"			
(TH) Thesaurus (syn/antonym)		Find synonyms for "frustration" in Macmillan and note them down. Different meanings of the word will take you to different groups of synonyms organized under a common idea or concept.		
(R) (DE) Related words/expressions Derived words	"return" Macmillan: see Other Entries for this word.			
(TRANS) Find a better translation/word choice	Find translations for "encargarse" ii and see which one fits sentence 1			Find at least 2 ways to express "quedarse tranquilo" iii in English. Read and note down the examples and their translations in Cambridge or Merriam Webster bilingual dictionaries.
(i) lleno de frustración: in/with frustration; (ii) encargarse: see to; (iii) quedarse tranquilo: stay calm				

NOTES FOR RUBRIC COMPLETION: Students are required to (1) provide a correct version of the sentences, (2) find and note down the information required in the vocabulary expansion cells (for example find different ways to say something or gather synonyms, or provide new sentences with precise words). Students will get an active dictionary rubric every other week during one academic semester. Active dictionary rubric completion is followed by class feedback and correction.

3.   Participants

A group of 75 first-year ESP university students enrolled in the Criminology and Security degree participated in the study. The group had a B1 level (37 points in the 30–49 range) in the Common European Framework of Reference (CEFR) as reported after completing the Oxford Placement Test. Participants live in a bilingual area where Spanish and Valencian are spoken. Students coming from other areas in the country as well as immigrants are taught both languages at elementary school and secondary school.

4.   Procedure

A pre-training questionnaire elicited information on dictionary use and background knowledge. Micro-writing tasks (see first row in Table 2 as an example) were then set out in which the students' writings were used to design the active dictionary rubric. Since the rubric was intended to help them with their own writings, this was considered as a useful practice in the sense that students could become aware of their progress in using dictionaries and how dictionaries could help them write better. A second post-training questionnaire interrogated the students on the training experience in order to compare results with their initial dictionary use and dictionary skills habits. Both questionnaires can be seen in the Appendix.

5.   Results

5.1   Pre-training questionnaire

A questionnaire of 8 questions (Q1 to Q8), six quantitative questions and two qualitative ones, was answered at the beginning of the course using Moodle. Table 3 illustrates the age and first dictionary type that the students employed. As can be seen in the table, the majority of students used a dictionary for the first time when they were between 8 and 10 years old, followed by a lower age range between 11 and 14 years old. Thirty-seven students started with a monolingual Spanish dictionary, followed by English–Spanish dictionaries, and Spanish–Catalan ones. It is remarkable not only that less than 50% started with a monolingual mother tongue dictionary but also that the second place was taken by the bilingual English–Spanish dictionary considering that the majority of students used this dictionary at an early age.

I used my first dictionary when I was …		The dictionary was …	
8–10 years old	63 (84.00 %)	A Catalan monolingual dictionary	3 (4.00 %)
11–14 years old	11 (14.67 %)	A Spanish monolingual dictionary	37 (49.33 %)
15–16 years old	0	A Spanish/Catalan Bilingual dictionary	14 (18.67 %)
16–18 years old	0	An English–Spanish bilingual dictionary	20 (26.67 %)
older than 18	1 (1.33 %)	An English monolingual dictionary	1 (1.33 %)

Table 3:   Q1 and Q2: Age and dictionary type for participants' first encounter with dictionaries

The following question, (Q3) "Has anyone ever taught you how to use a dictionary (explanation longer than 10 minutes)? If so, please explain who, how and when" aimed at eliciting any experience of practical dictionary training. It was considered that a 10 minutes single explanation would not count as proper training. It was found out that 23 out of 75 students never got any training at all some simply said that they were not taught and a few answered that no one had taught them and that they had taught themselves. Those who gave a positive answer explained that they were taught to find entry words by first finding the first letter of a word and then the full word. Only two students mentioned definitions, only saying that teachers helped them find where the definition was. No other mention was made of word meaning or word senses or any other dictionary component. Some students stated that they needed dictionaries because they did not know the language(s). Both teachers and parents appear in the students' answers as they explained how to use dictionaries. Student sample answers are:

— my mom has explained me in our house quite good for ten years but I didn't listen her
— No, they just said me to use an application for the mobile phone which offers me an online dictionary
— No, I know how to use a English dictionary because is similar to Spanish version

— My teacher in primary school taught me to use it, she explained me how to search the words.
— Yes, my teacher when I was in the primary school. She explain to us to search the words watching the word that appears in the top and then think if our word stay before or after in alphabetical order.
— My English teacher when I came to Spain and I didn't know how to even talk in Spanish. So, she had to explain to me with a dictionary on her hand.
— My parents when I was a child, each word that I don't understand they obligue me search it in a dictionary. they explain me how do it, i have got know the first letter of a word that I want to find. the dictionary are ordened by abecedary, then i had find the letter that I was searching following the abecedary.
— No one taught me how to use it, I informed myself.
— generally nobody has taught me how to use the dictionary but sometimes in secondary school teachers have given me links to dictionaries

From the students' comments, it was revealed that the explanations received and class practice seem to be very basic for the participants in this study and do not seem to cover any skill beyond locating information. When questioning students on their satisfaction regarding their ability to use dictionaries, however, the majority of students were satisfied and only 1 of them was not happy with his/her dictionary skills and another student stated that he/she would like to improve such skills.

Are you happy with your abilities to use dictionaries?	Responses
Yes, I'm totally satisfied	18 (24.00 %)
Yes, I manage well	43 (57.33 %)
Yes, more or less	10 (13.33 %)
No, not really	1 (1.33 %)
I would like to use it better	1 (1.33 %)
No answer	2 (2.67 %)

Table 4:   Q4: Students' perception of their skills in dictionary use

As regards the use of online dictionaries (Table 5), it was found out that most students use between two and three online dictionaries instead of sticking to a favourite dictionary, a practice followed by only 16% of the students. Of these, students were questioned on the type of dictionary they selected and for this question they were allowed more than one answer. Results indicated that 80.82%

choose bilingual dictionaries, closely followed by thesauri. Although mono-lingual, bilingual and bilingualised dictionaries may be both general or spe-cialized, these last two options were included to find out about the students' needs regarding terminology in a particular discipline or the lack of it. The results were as expected of this particular group of students. Since they have just started their degree in Criminology and Security some of them (13.7%) start to consult information on specialised terms while the majority are still looking for general vocabulary.

How many different online dictionaries do you usually look up?		Please tick the dictionary types you use (more than one answer is allowed)	
1	12 (16.00 %)	general	32 (43.84 %)
2	33 (44.00 %)	specialized	10 (13.70 %)
3	20 (26.67 %)	bilingual	59 (80.82 %)
4	4 (5.33 %)	monolingual	30 (41.10 %)
more than 4	2 (2.67 %)	thesaurus (synonyms, antonyms)	34 (46.58 %)
		bilingualised (mono-lingual and ALSO the translation of the word, sometimes the transla-tion of the examples)	26 (35.62 %)

Table 5:   Q5 and Q6: Number of different online dictionaries the participants use and types of most frequently used dictionaries

The following question was Q7 "What do you use your dictionaries for NOW? If you use them for different purposes, please explain." There were three recur-rent purposes: finding a translation, word meaning and finding synonyms. Some students also highlight a more general purpose: learning a language. While most students explain the purpose of dictionary use in general, a few students specify the type of dictionary they use depending on their needs. Sample answers are shown below:

— Now I use thesaurus dictionary when I not want repeat the same word in a writings all the time. I use the general dictionary to look for the meaning of words when I'm studing. Finally I use the bilingualised dictionary to look for the traduction of words in other languages as English when I write writings or I study it.
— to check vocabulary
— I use them mainly for translate words from spanish to english, but I also use it when I'm not sure of the meaning of a word or an specific sentence.
— I use them to look at specificated words that i don't remember how it were written
— I use the dictionary to translate into Spanish words that I do not understand in English and I also use it for university work
— I use dictionaries to search a word that I didn't understand.
— to be able to complete sentences

Students were also asked if they used those sections in the dictionary that explain how to use it (Table 6). The majority of students did not use that type of information. This is also a question to consider in relation to dictionary training: if students are not taught and do not read any guidance matter, then it is difficult that they come to appreciate dictionaries and their pedagogical innovations. The percentage of students who used the dictionary guidance was even lower (8%) than the findings reported by Lew and Galas (2008) on the same item (four students (14.3%) in the experimental and six (20.7%) in the control group). If we bear in mind that students in the present study were university students and those in Lew and Galas (2008) were primary school students, this can be seen as a major flaw in this particular group of Spanish university students.

Do you read the "how to use the dictionary" sections?	Responses
yes	6 (8.00 %)
no	52 (69.33 %)
sometimes	13 (17.33 %)
no answer	4 (5.33 %)

Table 6:   Q8: Using "how to use the dictionary" sections in dictionaries

5.2   Training sessions

During the first session, one micro-writing task was assigned. This task was used to generate the first two active dictionary rubrics. In the following train-

ing session, one active dictionary rubric was explained with examples on how to use it, and the students were given another active dictionary rubric for homework. In the following sessions, one micro-writing task was assigned, the last active dictionary rubric was revised and a new active dictionary rubric was given. Revision of active dictionary rubrics entailed that students had to present their corrected sentences after following the rubric, as well as the notes they took as instructed in each rubric. During the training sessions, it was essential to solve any question regarding dictionary use for that particular task. Projecting the teacher's screen to show the specific entries that were consulted for the task aided in task comprehension and as part of the process of dictionary skills training.

5.3   Post-training questionnaire

After the training period, students were asked to complete a second questionnaire with five questions, three quantitative and two qualitative questions. The main aim of this questionnaire was to analyse the effect of dictionary training on the students.

The first question asked students whether they managed to find information in the dictionaries. This was an important question because it provides feedback on whether the rubric could be understood and whether it easily took the learner to the right places. Positive results confirm the usefulness of this type of rubric: 49.33% of the students said they found the information, as well as another 24% who stated that they nearly always found what they were looking for. The percentage of students with difficulties was very low.

Did you manage to find the information in the dictionaries?	Responses
yes	37 (49.33 %)
nearly always	18 (24.00 %)
sometimes	12 (16.00 %)
it was difficult but I found it	7 (9.33 %)
no	1 (1.33 %)

Table 7:   Post-training Q1 for ease of information finding

The second question was about the usefulness of rubrics. The answers to this question indicated that participants found the practice either useful or more or less useful. Only three students did not find the rubric particularly useful.

Was the dictionary rubric useful for you?	Responses
yes	35 (46.67 %)
more or less useful	35 (46.67 %)
I do not know	2 (2.67 %)
no	3 (4.00 %)

Table 8:   Post-training Q2 Rubric usefulness

Question 3 is related to the previous question, but was posed differently. While question 2 is general, question 3 is a little bit more specific in the usefulness of the rubric by asking students whether locating information in the dictionaries is now easier than before the training.

Do you now find more information in dictionaries than before the rubric exercises? (more than one answer is allowed)	Responses
Yes, now I find many more things that I didn't know before	29 (38.67 %)
Yes, now I know how to learn more things about the words I am looking for	37 (49.33 %)
Yes, now I can do things with the dictionaries that I had never done before	5 (6.67 %)
Maybe I find more things than before this experience	20 (26.67 %)
Now I find one new type of information, I didn't know it existed before the rubric exercises	9 (12.00 %)
No, I do not find more information than before	2 (2.67 %)

Table 9:   Post-training Q3: finding information in dictionaries

Answers in table 9 indicated that the majority of students found it easier to both locate different types of information that they would not find or use prior to this study, and that they are now able to teach themselves new aspects of the words they might be interested in their future situations of dictionary usage.

The fourth question, (Q4) "Do you think your abilities to use dictionaries are now better? Why?" is related to the perception students had of their own dictionary skills after the training experience. Only five students gave a negative answer to this question, while the remaining seventy students affirmed that their dictionary skills had improved. Participants' answers point to the improvement in their ability to find different information types in a more effective way. Many of the answers also stated that they did not use dictionar-

ies very much before the training and that now they see the point in using them. Other aspects of dictionary use that appeared in the answers to this question were the students' ability to make a better word choice, paying attention to the context of examples, the value of creating a habit or familiarizing themselves with these tools. Some sample answers from students are shown below:

— Yes, because now I can find more things that before I don't see or that I don't give value, and now I'm faster to find the information.
— Yes, I think my abilities to use dictionaries is now better, becouse I can find new useful information for improve my redactions* such as, synonyms or the correct word depending the context. (*a mistranslation from Spanish "redacción", meaning "essay")
— I think that they had improved a little bit more than before because now I search in a more specific way to look for what I want more quickly.
— Yes, because I've learned a lot with the rubrics and now I've discovered lots of dictionaries I never used before and how to find the information in them.
— Yes, I think that my capacities have improved because before it was not using very much the dictionaries, however, now my vocabulary is richer and varied.
— I think that my abilities are better now but not too much.
— yes, because I used them so many times and I familiarized with them.
— Yes, I think my abilities to use dictionaries are better because now I have the habit of doing it, before I almost didn't do it.
— No, but I know which dictionaries I can use now.
— Yes, now I am able to contrast information in order to know which word is better for each situation
— In my opinion I believe that my skills with dictionaries are better, before I only looked for what the word meant, but now I look for if the word has some particle, or how to put the following verbs ...
— Yes, because when you had the indications to find the information you got some skills to search in the dictionaries.

The last question of the second questionnaire was Q5 "Tell me the things you can now do with dictionaries that you never did before". Answers to this question revealed how the majority of students are now more aware of the value of dictionaries as part of their language learning process. While most students said that they used dictionaries for two specific purposes, looking up the meaning of a word and translating the word, and for the general purpose of learning a language, they reported a number of newly gained insights and purposes in their post-training answers to Q5. Among these are the usefulness of example sentences for guidance on word usage, locating and learning about phraseology and idioms related to a word, distinguishing between different word senses, or choosing the right word among several options. Some students

also pointed to the fact that their dictionary use was now less time-consuming. Interestingly, dictionary training also proved to be a good instrument in raising students' awareness on the added values of using dictionaries in contrast to the use of Google translator, because they realised that dictionaries enable them to make informed decisions on their word choice and use.

— Use the examples to put in a better way the words inside a sentence and to use the prepositions that go before.
— For example, now I can find expressions in dictionaries and before I didn't know that it was possible
— Now, I have learned that some dictionaries like Cambridge or Macmillan are more useful than Google traductor.
— Before, I didn't search information in English dictionaries. Now, I can look words and meanings without needing the translator.
— For example, in the Cambridge dictionary you can see the grammar of words.
— Contrast information, know which preposition follows each word, read many examples...
— I have discovered the different meanings of a word
— When I'm reading a definition in a dictionary, now I can distinguish the type of word and possible expressions with the word.
— I can know if it is better to use another verb than the one I had chosen. I can know if that word is followed by a particle in particular.
— Before I read the definition of the words in English, I only looked for the meaning in Spanish and associated it with what I thought in Spanish. Now I also read in what situations the word can be used.
— I can find the words more quickly
— Now I can find phrases with the words I'm looking for for a better understanding of their meanings, expressions with those words and the most successful synonyms

6.   Discussion

The comparison of students' skills and beliefs in their own dictionary use revealed important changes between pre- and post-training opinions. As reported in the pre-training questionnaire results, most students received very basic instructions on how to use dictionaries and these instructions are provided at a very early age (8–10 years old) for the participants' mother tongue dictionaries or the official languages spoken in the area. It seems that after that, no additional training is given other than suggesting suitable dictionaries to use in a foreign language. This is a situation which could be improved by implementing different dictionary skills training progressively from basic to more difficult abilities in the different educational levels. In spite of the little training, students believed themselves to be proficient dictionary users and 61.33% claimed

to be satisfied or manage well with dictionaries. This perception, however, changed after the training sessions. In the second questionnaire students stated that the active dictionary rubrics were useful (70 out of 75 students answered that they were useful or more or less useful), thus indicating that there was room for improvement, which can be achieved by instruction and training. This idea was further supported by the answers provided to the question "Do you now find more information in dictionaries now than before the rubric experience?" For this question, the majority of students either found more things that they did not know before or had learnt how to learn more things about words. Moreover, students reported that their abilities to use dictionaries was better after the training period.

Questionnaire results also indicated that the students' skills now go beyond the three typical dictionary use options (translation, definition (word meaning), and a general sense that dictionaries may aid in language learning). Dictionary use and training better acquaint students with dictionaries, offering a wider range of linguistic information that may help and make the students more confident in their language learning process.

7.   Conclusions and pedagogical implications

This study is in line with previous research (literature review, this article) on the value of dictionary skills for the enhancement of language learners' linguistic competence. Conclusive results in this and previous studies on the improvement of learners' linguistic competence after dictionary instruction suggest that dictionary skills training is an objective that should be pursued by language teachers if they want to provide their learners with useful, adequate language tools that can also become life-long learning companions.

One of the findings of this and previous studies (particularly, Lew and Galas 2008) is the fact that most learners believe that they know how to use dictionaries and that there is not much that they can learn about them. However, training in dictionary skills shows them that they were missing a considerable amount of information that could help them in their learning process and that goes beyond translating a word or being able to find one definition for a word. Students' answers acknowledge that dictionaries allow them to better find their way with language use, helping them make informed decisions on word choice while learning about word usage. This acknowledgement includes awareness of the difference between obtaining a translation in Google or other online translation tools and deciding which is the best way to say what they want to say. Thus, specific training methods of dictionary users need to be part of the equation when reporting users' abilities to use dictionaries for a specific purpose in a given context.

Some pedagogical implications can be derived from studies in dictionary training. A recurrent conclusion is that consistent and systematic practice is necessary to achieve dictionary skills competences at different levels. Another

aspect that calls for closer attention is that the use and training reported in the literature refers to different dictionaries and types of learners. Further research should systematise these two aspects of research in dictionary skills, since both dictionary complexity and language learner competency level are relevant parts of the analysis. In this line of thought, researchers should consider how practice with different (online) dictionaries and the affordances each may have, could yield different results in the students' performance.

Finally, new methods to approach dictionary skill training and material provision needs to be aligned with language competence frameworks (Campoy-Cubillo 2015) so that dictionary training at different educational stages can be carefully planned out.

8.   Bibliography

8.1   Online dictionaries

Cambridge Dictionary Online. https://dictionary.cambridge.org/.
Macmillan English Dictionary Online. http://www.macmillandictionary.com.
Merriam-Webster Dictionary Online. http://www.merriam-webster.com.
Merriam-Webster Spanish Central Online. https://www.spanishcentral.com/.

8.2   Other references

Ali, N. Al Amin. 2017. Efficacy of Explicit Teaching of Lexicographical Skills: The Case of Sudanese English Majors. *International Journal of Research in English Education* 2(3): 47-66.

Allan, D. 2001. *Oxford Placement Test.* Oxford University Press and University of Cambridge Local Examinations Syndicate (UCLES).

Bishop, G. 2001. Using Quality and Accuracy Ratings to Quantify the Value Added of a Dictionary Skills Training Course. *The Language Learning Journal* 24(1): 62-69.

Bogaards, P. 2003. Uses and Users of Dictionaries. Van Sterkenburg, Piet. (Ed.). 2003. *A Practical Guide to Lexicography:* 26-33. Amsterdam/Philadelphia: John Benjamins.

Campoy-Cubillo, M.C. 2015. Assessing Dictionary Skills. *Lexicography* 2: 119-141.

Carduner, J. 2003. Productive Dictionary Skills Training: What Do Language Learners Find Useful? *The Language Learning Journal* 28(1): 70-76. DOI: 10.1080/09571730385200231.

Chi, M.L.A. 2003. *A Study of Dictionary Use Teaching.* PhD Dissertation. Sydney: Macquarie University.

Chi, M.L.A. 2020. Reconstructing the Lexicographical Triangle through Teaching Dictionary Literacy to Teachers of English. *Lexicography: Journal of ASIALEX* 7(1–2): 79-95. https://doi.org/10.1007/s40607-020-00066-x.

Dziemianko, A. 2016. An Insight into the Visual Presentation of Signposts in English Learners' Dictionaries Online. *International Journal of Lexicography* 29(4): 490-524.

Dziemianko, A. 2019. Homogeneous or Heterogeneous? Insights into Signposts in Learners' Dictionaries. *International Journal of Lexicography* 32(4): 432-457.

Egido Vicente, M. 2021. Dictionaries in German and Spanish Primary Education Curricula: A Comparative Study. *International Journal of Lexicography*; ecab020, https://doi.org/10.1093/ijl/ecab020.

Hadebe, S. 2004. Improving Dictionary Skills in Ndebele. *Lexikos* 14: 89-104.

Laufer, B. 2011. The Contribution of Dictionary Use to the Production and Retention of Collocations in a Second Language. *International Journal of Lexicography* 24: 29-49.

Lew, R. and K. Galas. 2008. Can Dictionary Skills be Taught? The Effectiveness of Lexicographic Training for Primary-school-level Polish Learners of English. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008:* 1273-1285. Barcelona: Universitat Pompeu Fabra.

Liang, P. and D. Xu. 2018. An Empirical Study of EFL Learners' Dictionary Use in Chinese–English Translation. *Lexikos* 28: 221-244.

Liu, Lin. 2014. The Integration of Dictionary Use Strategy Training into Basic English Class. *Theory and Practice in Language Studies* 4(10): 2138-2143. DOI:10.4304/tpls.4.10.2138-2143.

Lopera, S.A. 2019. Effects of Reading Strategy and Dictionary Instruction in an Undergraduate Foreign Language Reading Comprehension Group. *Folios* 50: 127-138. doi: 10.17227/Folios.50-10226.

Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries.* (Lexicographica Series Maior 145.) Berlin/Boston: De Gruyter.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nkomo, D. 2015. Developing a Dictionary Culture through Integrated Dictionary Pedagogy in the Outer Texts of South African School Dictionaries: The Case of 'Oxford Bilingual School Dictionary: IsiXhosa and English'. *Lexicography: Journal of ASIALEX* 2(1): 71-99. https://doi.org/10.1007/s40607-015-0021-8.

Ptasznik, B. 2018. Types of Sense-navigation Devices in Print Monolingual English Learners' Dictionaries. *Prace Językoznawcze* 20(2): 157-170.

Ptasznik, B. and R. Lew. 2019. New-line and Run-on Guiding Devices in Print Monolingual Dictionaries for Learners of English. *Lexikos* 29: 180-198.

Ranalli, J. 2013a. Online Strategy Instruction for Integrating Dictionary Skills and Language Awareness. *Language Learning & Technology* 17(2): 75-99.

Ranalli, J. 2013b. Designing Online Strategy Instruction for Integrated Vocabulary Depth of Knowledge and Web-based Dictionary Skills. *CALICO Journal* 30(1): 16-43.

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

Schofield, P. 1999. Dictionary Use in Reception. *International Journal of Lexicography* 12(1): 13-34.

Tono, Y. 1989. Can a Dictionary Help One Read Better? On the Relationship Between EFL Learners' Dictionary Reference Skills and Reading Comprehension. James, G. (Ed.). 1989. *Lexicographers and Their Works:* 192-200. Exeter: University of Exeter Press.

Yorkey, R.C. 1982. *Study Skills for Students of English as a Second Language.* New York: McGraw-Hill.

Appendix

Active dictionary rubric

Active dictionary rubric	S 1	S 2	S 3	S 4	S 5
(DEF) Definition: check definition for a better understanding/use of word					
(SENS) Check you have chosen the correct sense of the word					
(COL) Check, learn or use correct collocation					
(EX) Check examples for a more natural use of word(s) / expressions (click on "more examples" if you see this link in the dictionary)					
(GRMM) Read the grammar information about this word					
(US) Usage / get it right					
(IDM) Look at the "More meanings" tab to find idioms, phrases and phrasal verbs related to this word to learn new expressions					
(TH) Thesaurus (synonym/antonym)					
(R) (DE) Related words/ expressions / Derived words					
(TRANS) Find a better translation/word choice					

Pre-training questionnaire

(Q1) I used my first dictionary when I was …
- ☐ 8–10 years old
- ☐ 11–14 years old
- ☐ 15–16 years old
- ☐ 16–18 years old
- ☐ older than 18

(Q2) The dictionary was …
- ☐ A Catalan monolingual dictionary
- ☐ A Spanish monolingual dictionary
- ☐ A Spanish/Catalan Bilingual dictionary
- ☐ An English–Spanish bilingual dictionary
- ☐ An English monolingual dictionary

(Q3) Has anyone ever taught you how to use a dictionary (explanation longer than 10 minutes)? If so, please explain who, how and when.

(Q4) Are you happy with your abilities to use dictionaries?
- ☐ Yes, I'm totally satisfied
- ☐ Yes, I manage well
- ☐ Yes, more or less
- ☐ No, not really
- ☐ I would like to use it better
- ☐ No answer

(Q5) How many different online dictionaries do you usually look up?
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ More than 4

(Q6) Please tick the dictionary types you use (more than one answer is allowed).
- ☐ general
- ☐ specialized
- ☐ bilingual
- ☐ monolingual
- ☐ thesaurus (synonyms, antonyms)
- ☐ bilingualised (monolingual and ALSO the translation of the word, sometimes the translation of the examples)

(Q7) What do you use your dictionaries for NOW? If you use them for different purposes, please explain.

(Q8) Do you read the "how to use the dictionary" sections?
- ☐ yes
- ☐ no
- ☐ sometimes
- ☐ no answer

Post-training questionnaire

(Q1) Did you manage to find the information in the dictionaries?
- ☐ yes
- ☐ nearly always
- ☐ sometimes
- ☐ it was difficult but I found it
- ☐ no

(Q2)   Was the dictionary rubric useful for you?
- ☐   Yes
- ☐   More or less useful
- ☐   I do not know
- ☐   No

(Q3)   Do you now find more information in dictionaries now than before the rubric experience? (more than one answer is possible)
- ☐   Yes, now I find many more things that I didn't know before
- ☐   Yes, now I know how to learn more things about the words
- ☐   I am looking for
- ☐   Yes, now I can do things with the dictionaries that I had never
- ☐   done before
- ☐   Maybe I find more things than before this experience
- ☐   Now I find one new type of information, I didn't know
- ☐   it existed before the rubric exercises
- ☐   No, I do not find more information than before

(Q4)   Do you think your abilities to use dictionaries is now better? Why?

(Q5)   Tell me the things you can now do with dictionaries that you never did before.

Beryl T. (Sue) Atkins
Leksikograaf van A tot Z

Dit is onmoontlik om 'n omvattende oorsig te gee van wat die legendariese Sue Atkins tot wêreldleksikografie en die Suid-Afrikaanse leksikografie in die besonder bygedra het. Soos genoem in die voorwoord van 'n vorige *Festschrift in Honour of B.T.S. Atkins*, was sy 'n inspirasie vir generasies leksikograwe en het sy 'n deurslaggewende rol in die ontwikkeling van skakels tussen die praktiese wêreld van woordeboekmaak en die navorsingsterreine van die linguistiek en natuurliketaalprosessering gespeel. Sy was alombekend vir haar praktiese en pragmatiese benadering tot woordeboekmaak, haar inherente linguistiese intuïsie en aanvoeling vir die leksikografie wat deur jare se ervaring in die samestelling van woordeboeke gevorm en geslyp is. Sue Atkins was sowaar iemand met 'n groot passie en onuitputlike entoesiasme vir die leksikografie. Onder haar talle prestasies is dit waarskynlik haar leidende rol in die konseptualisering van die Britse Nasionale korpus waarvoor sy onthou sal word — die Britse leksikografie is gerevolusioneer deur die beskikbaarheid van die korpus en korpusgebaseerde leksikografie het deel geword van die standaard leksikografiese praktyk in baie dele van die wêreld, o.a. in Suid-Afrika. Sy het talle professionele en akademiese opleidingskursusse en werkswinkels in die leksikografie georganiseer waar sy ook onderrig het, en het leksikografiese konsultasiewerk in oorleg met 'n aantal toonaangewende instansies onderneem. Sy het meegedoen aan belangrike nasionale en internasionale navorsingsprojekte in die rekenaarleksikografieveld.

Haar talle eerbetone en akademiese toekennings, waarvan slegs 'n paar hier gelys word, dien as bewys van haar besondere status as internasionale figuur in die wêreld van die leksikografie. In 2000 het sy 'n eredoktorsgraad van die Universiteit van Brighton, VK, ontvang, gevolg deur 'n festschrift *Lexicography and Natural Language Processing* in 2002, wat deur EURALEX gepubli-

seer is ter viering van haar bydrae tot die internasionale leksikografie. Sy is ook verklaar tot lewenslange erelid van EURALEX, waarvan sy voorheen president was. In 2008 is 'n eredoktorsgraad deur die Universiteit van Pretoria aan haar toegeken vir haar uitsonderlike bydraes tot die leksikografiese praktyk wêreld-wyd, maar spesifiek ook vir die rol wat sy gespeel het in die vestiging van 'n goeie leksikografiese praktyk vir die Afrikatale in Suid-Afrika.

Haar direkte betrokkenheid by die Afrikataalleksikografie het 'n aanvang geneem met die SALEX-tutoriaal van 1997, wat by die Rhodes Universiteit, Grahamstad, gehou is waartydens sy 'n weeklange opleidingskursus in die samestelling van eentalige woordeboeke aangebied het. Dit is opgevolg deur AFRILEX-SALEX '98, 'n tweeweeklange opleidingskursus in die samestelling van tweetalige woordeboeke wat by die Universiteit van Pretoria aangebied is. As gevolg van hierdie twee grondverskuiwende werkswinkels en die daarop-volgende wye belangstelling in die ontwerp en skep van woordeboeke vir hier-die tale, het woordeboekmaak in Suid-Afrika verander van lukrake, onsistema-tiese publikasies na 'n ernstige toewyding aan leksikografiese beginsels en praktyk. Al hoe meer woordeboeke van hoogstaande gehalte word nou in Suid-Afrika geskep.

Sue Atkins sal onthou word vir haar toewyding aan die leksikografie, maar ook as 'n deernisvolle mens met 'n egte belangstelling in ander en ook in die omgewing, asook vir haar alomteenwoordige sin vir humor.

D.J. Prinsloo
Departement Afrikatale
Universiteit van Pretoria
Suid-Afrika
(danie.prinsloo@up.ac.za)

Beryl T. (Sue) Atkins
Lexicographer from A to Z

It is not possible to give a comprehensive overview of what the great Sue Atkins has done for lexicography in the world and South African lexicography in particular. As stated in the foreword to a previous *Festschrift in Honour of B.T.S. Atkins*, she was an inspiration to generations of lexicographers and played a crucial role in developing links between the practical world of dictionary-making and the research communities of linguistics and natural language processing. She was well-known for her practical and pragmatic approach to dictionary making, her inherent linguistic intuition and feel for lexicography shaped and sharpened by years of experience in dictionary compilation. If ever there was someone who had a passion and unfailing enthusiasm for lexicography, it was Sue Atkins. Amongst her many accomplishments, it is probably her leading role in the conceptualization of the British National corpus that she will be remembered for — British lexicography was revolutionized by the availability of the corpus and corpus-based lexicography has become part of standard lexicographic practice in many parts of the world, South African being one of them. She has organized and taught at many professional and academic training courses and workshops in lexicography, and held lexicographic consultancies with a number of prestigious research institutions. She has participated in important national and international research projects in the field of computational lexicography.

Her exceptional standing as international figure in the world of lexicography is evidenced by her many accolades and academic awards, of which only a few are listed here. In 2000, she received an Honorary DLitt from the University of Brighton, UK, followed by a festschrift *Lexicography and Natural Language Processing* in 2002, published by EURALEX to mark her contribution to international lexicography. She was also made an Honorary Life Member of EURA-

LEX, of which she is a past President. In 2008 she was awarded an honorary doctorate by the University of Pretoria for her significant contributions to lexicographic practice worldwide, but also specifically for the role she played in the establishment of good lexicographic practice for the African languages in South Africa.

Her direct involvement in African language lexicography started with the SALEX Tutorial of 1997, held at Rhodes University, Grahamstown where she presented a one-week training course on the compilation of monolingual dictionaries. This was followed up by AFRILEX-SALEX '98, a two-week training course on the compilation of bilingual dictionaries offered at the University of Pretoria. As a result of these two ground-breaking workshops and the subsequent widespread interest in designing and producing dictionaries for these languages, dictionary-making in South Africa has shifted from random, haphazard publications to a serious commitment to lexicographic principles and practice. More and more dictionaries of a high quality are now being produced in South Africa.

Sue Atkins will be remembered for her commitment to lexicography, but also as a compassionate human being, taking a genuine interest in others and also in the environment, and for her ever-present sense of humour.

D.J. Prinsloo
Department of African Languages
University of Pretoria
South Africa
(danie.prinsloo@up.ac.za)

Publikasieaankondigings / Publication Announcements

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Sestiende Deel: SRP–SZONDITOETS.* 2021, xx + 759 pp. ISBN-13 978-1-990998- 46-1 (leerband), ISBN-13 978-1-990998-45-4 (plastiekband). Stellenbosch: Buro van die WAT. Prys: R900 (leerband) / R550 (plastiekband). (Resensie in hierdie nommer.)

Alex Duffey en Susan Roets (Reds.). *Kunswoordeboek / Arts Dictionary.* 2021, 392 pp. ISBN: 9781485300564 (sagteband). Pretoria: Protea Boekhuis. Prys: R495.

M.M.M. Duka. *Shuters IsiChazi-magama SesiXhosa*. 2018, xvi + 237 pp. ISBN 978-0796058218 (Hardback). Pietermaritzburg: Shuter and Shooter. Price: R104.65 (Hardback). (Review in this issue.)

Mariusz Piotr Kamiński. *Defining with Simple Vocabulary in English Dictionaries.* 2021, xv + 326 pp. ISBN: 9789027208590 (Hardback), ISBN: 978-9027260000 (e-Book). Terminology and Lexicography Research and Practice. Volume 22. Amsterdam/Philadelphia: John Benjamins. Price: €99.00 (Hardback and e-Book).

Peter Martin. *The Dictionary Wars: The American Fight over the English Language*. 2019, xii + 358 pp. ISBN 978-0-691-18891-1. Princeton: Princeton University Press. Price: $14.61 (Paperback) / $29.95 (Hardback). (Review in this issue.)

Sarah Ogilvie and Gabriella Safran (Eds.). *The Whole World in a Book: Dictionaries in the Nineteenth Century.* 2020, 358 pp. ISBN: 978-0190913199 (Hardback). Oxford: Oxford University Press. Price: £22.99. (Review in this issue.)

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and Afrikaans*. Second edition. 2021, 116 pp. ISBN 9780190758219 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and isiXhosa*. Second edition. 2021, 116 pp. ISBN 9780190758226 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and isiZulu*. Second edition. 2021, 116 pp. ISBN 9780190758233 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and Sesotho*. Second edition. 2021, 116 pp. ISBN 9780190758257 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and Sesotho sa Leboa*. Second edition. 2021, 116 pp. ISBN 9780190758240 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Oxford University Press Southern Africa. *Oxford First Bilingual Dictionary: English and Setswana*. Second edition. 2021, 116 pp. ISBN 9780190758264 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R149.95.

Pharos. *Pharos Bilingual English–isiNdebele / isiNdebele–English Dictionary*. 2021, 192 pp. ISBN: 9781868902507 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–IsiXhosa / IsiXhosa–English*. 2021, 192 pp. ISBN: 9781868902439 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–isiZulu / isiZulu–English*. 2021, 192 pp. ISBN: 9781868902422 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–Sepedi / Sepedi–English*. 2021, 192 pp. ISBN: 9781868902453 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual English–Sesotho / Sesotho–English Dictionary*. 2021, 192 pp. ISBN: 9781868902446 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–Setswana / Setswana–English*. 2021, 192 pp. ISBN: 9781868902460 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–SiSwati / SiSwati–English*. 2021, 192 pp. ISBN: 9781868902484 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual English–Tshivenḓa / Tshivenḓa–English Dictionary*. 2021, 192 pp. ISBN: 9781868902491 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos Bilingual Dictionary English–Xitsonga/Xitsonga–English*. 2021, 192 pp. ISBN: 9781868902477 (Paperback). Cape Town: Pharos Dictionaries. Price: R149.

Pharos. *Pharos grondslagfasewoordeboek / Foundation Phase Dictionary Afrikaans–Engels / English–Afrikaans*. 2021, xi + 226 pp. ISBN: 978-1868902408 (sagteband). Kaapstad: Pharos Woordeboeke. Prys: R220.

518

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: http://lexikos.journals.ac.za /)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels
Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, terminologie, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evaluerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Leksikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure
Lexikos is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index* (ISI).

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

3. Taal van bydraes
Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg
Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid
Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare
Lexikos is sedert volume 28 slegs elektronies beskikbaar op http://lexikos.journals.ac.za. Geen oordrukke of eksemplare is dus beskikbaar nie.

7. Uitnodiging en redaksionele adres
Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en verkieslik in elektroniese formaat aan die volgende adres te stuur: lexikos@sun.ac.za, of Die Redakteur: LEXIKOS, Buro van die WAT, Posbus 245, 7599 STELLENBOSCH, Republiek van Suid-Afrika.

B. VOORBEREIDING VAN MANUSKRIP
Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels
Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie **8 000 woorde** oorskry nie.

Elke artikel moet voorsien wees van 'n **opsomming** van ongeveer 200 woorde en ongeveer 10 **sleutelwoorde** in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde **in Engels**. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n **vertaling van die oorspronklike titel** bevat.

2. Grafika
Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks
Kyk na onlangse nommers van *Lexikos* vir meer inligting. Buiten in spesiale gevalle moet verwysings na *Lexikos*-artikels tot twee of drie per artikel beperk word. Uitsonderings moet met die redakteur van *Lexikos* uitgeklaar word. Dít word gedoen om die status van *Lexikos* in verskeie internasionale indekse te behou.

4. Aantekeninge/voetnote/eindnote
Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

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**.