An overview of Digital Lexicography and directions for its future: An interview with Gilles-Maurice de Schryver

Um panorama da Lexicografia Digital e rumos para o futuro: Uma entrevista com Gilles-Maurice de Schryver

Gilles-Maurice de Schryver
Ghent University
University of Pretoria
gillesmaurice.deschryver@ugent.be

Rove Chishman
Universidade do Vale do Rio dos Sinos
rove@unisinos.br

Bruna da Silva
Universidade do Vale do Rio dos Sinos
bbruna@edu.unisinos.br

Abstract: In this contribution Rove Chishman and Bruna da Silva present questions to the President of the *European Association for Lexicography*, Gilles-Maurice de Schryver. In answering, G-M de Schryver calls a spade a spade: No, there is no theory of lexicography, but yes, there is a future for lexicographers. For that future to be bright, lexicographers will have to join hands with the Big Data companies, and accept that their invaluable input will be all but invisible to the users. Artificial Intelligence components will have to be used in earnest, so that 'searches' (not 'look-ups'!) will instantly lead to context-sensitive answers. Browsing lexical data only starts from these answers and proceeds from there in two- (and why not three-) dimensional multimedia spaces. To achieve this, G-M de Schryver points to the most promising research ideas at present, as well as the top teams involved in digital dictionary user research.

Keywords: EURALEX; eLex; digital lexicography; e-dictionaries; corpora; automation; crowdsourcing; Corpus Query Packages (CQPs); Dictionary Writing Systems (DWSs); artificial intelligence (AI); user-friendliness; mobile devices; augmented writing assistants; eye-tracking; digital user research.

Resumo: Nesta contribuição, Rove Chishman e Bruna da Silva apresentam perguntas ao Presidente da *Associação Europeia de Lexicografia*, Gilles-Maurice de Schryver. Ao responder, G-M de Schryver vai direto ao ponto: Não, não há teoria da lexicografia, mas, sim, há um futuro para os lexicógrafos. Para que esse futuro seja brilhante, os lexicógrafos terão de tornar-se aliados das empresas que lidam com grandes quantidades de dados e aceitar que a contribuição inestimável que oferecem será praticamente invisível para os usuários. Os componentes da Inteligência Artificial terão de ser usados com seriedade para que as 'buscas' (como em ferramentas de busca; não as 'consultas', como em dicionários impressos) instantaneamente conduzam o usuário a respostas sensíveis ao contexto. A navegação por dados lexicais apenas começa com essas respostas e prossegue a partir daí por e para espaços multimídia bi- (e por que não tri-?) dimensionais. Para que isso seja alcançado, G-M de Schryver aponta as ideias de pesquisa mais promissoras atualmente, bem como as principais equipes envolvidas na pesquisa sobre usuários de dicionários digitais.

Palavras-chave: EURALEX; eLex; lexicografia digital; dicionários eletrônicos; corpora; automação; fontes de informação oriundas de massa; pacotes de consulta de grandes quantidades de textos; sistemas de escrita de dicionário; inteligência artificial; usabilidade; dispositivos móveis; assistentes de escrita aumentada; rastreamento ocular; pesquisa sobre usuário digital.

Interviewers

Dr Rove Chishman is a professor in the Applied Linguistics Graduate Program at Universidade do Vale do Rio dos Sinos (Brazil). She holds a PhD in Applied Linguistics from Pontifícia Universidade Católica do Rio Grande do Sul (Brazil) with a post-doctorate from the University of Texas at Austin (USA) under the supervision of Professor Hans Boas. Her research interests include the interface between Cognitive Semantics theories (with emphasis on Frame Semantics) and Lexicography, Corpus Linguistics and Computational Semantics. She is a full member of the Brazilian National Council of Research (CNPq). As the coordinator of the Research Group SemanTec – Semantics and Technology – professor Rove Chishman has managed to organize and publish online frame-based specialized dictionaries. Professor Rove currently coordinates two projects: the Paralympic Dictionary, which expands on the base of her previous work, and Semantic Technologies for Juridical Information Retrieval, which aims at developing resources for legal language.

Bruna da Silva holds a master's degree in Applied Linguistics from Universidade do Vale do Rio dos Sinos (Brazil) and is a PhD candidate (with funding by CAPES) at the same institution. She is a member of the research group SemanTec and has worked on the development of the Dicionário Olímpico. Currently, she conducts research for the Paralympic Dictionary. Her research interests include Cognitive Semantics, Digital Lexicography, Lexical Semantics and Corpus Linguistics. Her PhD dissertation being developed is concerned with drawing up proposals to study the current user behaviour of Dicionário Olímpico with an aim to further improve its efficacy.

Interviewed

Dr Gilles-Maurice de Schryver is a research professor of African linguistics in the department of languages and cultures at Ghent University (Belgium) and an extraordinary professor in the department of African languages at the University of Pretoria (South Africa). He holds an MSc in microelectronic engineering from University College Ghent, as well as an MA and PhD in African languages and cultures from Ghent University. As the author or co-author of over 300 books, book chapters, journal articles and conference papers, he has mainly worked on Bantu corpus linguistics and lexicography in general. His publications include award-winning dictionaries for Northern Sotho, Zulu and Xhosa, published with Oxford University Press, as well as various online dictionaries, including the most popular one for Swahili. He is the current President of EURALEX (2018-2020), the European Association for Lexicography, and a two-term past President of AFRILEX (2009-2013), its African sister association. He has also served in other capacities on the executive boards of AFRILEX (2001-2009), EURALEX (2006-2014), ASIALEX (2007-2013), for Asia, and AUSTRALEX (2008-2013), for Australasia.

Rove Chishman (RC) and Bruna da Silva (BS): With regard to your research area of African linguistics, how would you categorize the different types of research you have been developing with Bantu languages? Where do these studies differ from studies on languages, such as the English language, which have a much larger number of speakers around the world?

Gilles-Maurice de Schryver (G-MdS): For about two decades now, I have been working on and with various African languages, with a focus on the Bantu language family. There are well over 500 Bantu languages, spoken from the north of the Congo basin all the way to the south of the African continent. What makes them 'different' is not necessarily the fact that all of them have fewer speakers compared to English, but the fact that none of these languages has a written historical record. I will come back to this. Three research disciplines have caught my interest: lexicography, corpus linguistics, and natural language processing. Something about each now, in reverse order.

Together with Guy De Pauw and Peter Wagacha, we worked on part-of-speech (POS) taggers, diacritic restoration and machine translation for a number of African languages, most of them Bantu. Some of the tools we built have been brought together online at *African Language Technology* (De Pauw *et al.*, 2006-2019). The bottleneck here, compared to a big language like English, is that our languages are heavily under-resourced.

The earliest attempts to build corpora for Bantu languages date from the beginning of the 1990s, when D.J. Prinsloo (1991) began assembling material for Northern Sotho and Arvi Hurskainen (1992) did so for Swahili. Bantu corpus building gained momentum around the turn of the millennium, when I first built corpora for Cilubà in 1997-1998 followed by Swahili in 1999, and then joined forces with D.J. Prinsloo to eventually assemble corpora for all nine official Bantu languages of South Africa (De Schryver & Prinsloo, 2000). Now two decades later, corpora have been built for several dozen Bantu languages, ranging in size from a few million to several tens of millions of tokens each. All corpus-based and corpus-driven studies undertaken with those corpora are essen-

661

Figure 1. G-M de Schryver discusses evidence for prefix reduction in the Kikongo Language Cluster

[Photo by Heidi Goes, 4 May 2015]

The picture shows G-M de Schryver using evidence from the oldest Bantu dictionary — the 'Vocabularium Latinum, Hispanicum, e Congense', a manuscript from 1652 — to catch language change in the act for Kisikongo (spoken in Mbanza Kongo, Angola).

tially synchronic, owing to the fact that the Bantu languages have only been reduced to writing since a century or so ago. There is one set of Bantu languages for which we have attempted to go as far back in time as possible, namely for a number of varieties of the Kikongo Language Cluster (KLC). For the KLC variety known as Kisikongo, the oldest known document is a catechism published four hundred years ago (Cardoso, 1624) — which is also the oldest extant source for any Bantu language. Thanks to this and other sources from that period, namely a Latin-Spanish-Kisikongo dictionary manuscript (van Gheel, 1652) and a published Kisikongo grammar written in Latin (Brusciotto, 1659), as well as later sources for other KLC varieties (Descourvières, 1772; 1773; 1775; 1776), we have been able to complement the Comparative Method, common and very well developed in Bantu

linguistics, with evidence-based diachronic corpus-linguistic research, to make novel and highly original claims (Bostoen & De Schryver, 2015; 2018).

As to lexicography, across the African continent various missionaries, later followed by academics, approached the compilation of Bantu-language dictionaries in a number of ways, which led to two main lemmatisation traditions: word vs. stem (Van Wyk, 1995). Neither is necessarily better for all types of users, so when I began compiling school dictionaries for two Nguni languages from South Africa, I put the needs of the users first and broke with Nguni's so-called 'scientific' stem tradition; lemmatising orthographic words instead, basically going down the top section of corpus frequency lists. The dictionaries for Zulu and Xhosa produced in this way were published by Oxford University Press (De Schryver, 2010a; De Schryver & Reynolds, 2014), and have been very favourably received. Moreover, given that there was no satisfactory Dictionary Writing System available when we set out to compile our Bantu dictionaries, we developed our own lexicographic suite, known as TshwaneLex, or TLex for short (Joffe & De Schryver, 2002-2019), a piece of software which is now also used by hundreds of other dictionary teams the world over.

RC and BS: Every two years, eLex Conferences bring together Digital Lexicography researchers from around the world. Through an examination of the eLex 2019 theme description, one can notice expressions such as "user needs and habits", "dictionary use", and "dictionary users" appearing to be central to the description of conference goals. How would you describe the relationship between digital lexicography and research on dictionary use?

G-MdS: Research on dictionary use has a long history; indeed, it is at least a century old (Lew, 2011, p. 1). The names of Sue Atkins, Paul Bogaards, Reinhard Hartmann, Batia Laufer, Hilary Nesi and Yukio Tono immediately jump to mind. Some of the key publications in this field include Atkins (1998), Bogaards (1998), Hartmann (1987), Laufer & Melamed (1994), Nesi (2000) and Tono (2000). Caveat: There are more — Welker (2010) lists 320 dictionary-use studies — so these are actually just a few of the works which I enjoyed. Asking people what they want or do when they use dictionaries, or conducting small-scale tests using paper dictionaries is one thing, but being able to study digital dictionary use is a different thing altogether. The main reason is that it is more 'natural' to study dictionary use in an e-environment.

To begin with, in an e-environment a real, full dictionary may be used, for which the study can happen unobtrusively (De Schryver & Joffe, 2004). Typically, all user actions are logged into a 'log file'. Next, any number of parameters may be controlled in a fine-grained way, as may for instance be seen from the numerous excellent studies conducted by Anna Dziemianko; to name but her studies produced this year: Dziemianko (2019a; 2019b; 2019c). An e-environment furthermore allows to scale up, to include virtually any number of users, consulting any number of dictionaries, performing any number of searches: thousands, millions, billions of searches, logging data for years, even decades (De Schryver *et al.*, 2019).

While more digital usage data is often better data in this day and age of 'big data', such sizes may make it harder to control all parameters, and/or one may end up not knowing enough about who the dictionary users are. There will thus always be a need for small-scale, minutely-controlled, empirical dictionary-use studies. One particularly promising avenue of research that may be performed in connection with

digital dictionaries is the use of eye-tracking (Simonsen, 2009a; b; Kaneta, 2011; Simonsen, 2011; Tono, 2011; Lew *et al.*, 2013; Kemmer, 2014a; b; Müller-Spitzer *et al.*, 2014; Lew *et al.*, 2018).

RC and BS: From a dictionary-maker perspective, in what ways are digital dictionaries already able to serve users better than hard copy dictionaries? What really changed?

G-MdS: In one word: The use of large corpora during dictionary compilation, combined with software which helps to analyse and synthesise the huge amounts of data. From the perspective of the dictionary maker, this ensures being certain that nothing important has been overlooked; or even: that what is frequent and common is effectively covered in the dictionary, while knowing that one has safely omitted what is truly less important. The fact (or illusion) that there are no space constraints in a digital dictionary also helps to put all types of worries a lexicographer may have during compilation to rest. Examples of what changed now follow.

Take collocates: Long gone are the days that these had to be pinpointed manually by the lexicographer. Ever since the late Adam Kilgarriff introduced the concept and tool known as *Word Sketches* (Kilgarriff & Tugwell, 2001), viewing one-page summaries of how words typically behave in context has become a staple in Corpus Query Packages (CQPs).

Or take usage examples: Rather than having to limit supportive material, at best, to a few lines from the most respected authors (as typically found in traditional paper dictionaries) or, at worst, to a few concocted 'pedagogical examples' (Fox, 1987), lexicographers working towards a digital dictionary may safely illustrate all meaning potentials with 'real' examples, whereby the meanings are mapped onto the multitude of usages seen in their corpora (Hanks, 2002). In order to avoid information overload, good digital dictionaries allow their users to decide how many and for which senses examples are brought up. As to the lexicographers, ever since Adam Kilgarriff introduced the concept and tool known as GDEX – *Good Dictionary Examples* (Kilgarriff *et al.*, 2008), being able to evaluate sentences with respect to their suitability to serve as dictionary examples also became part of CQPs.

Combining *Word Sketches* with GDEX led to the concept and tool known as *TickBox Lexicography* (Kilgarriff *et al.*, 2010), whereby dictionary entry drafts are generated automatically. All of these tools (and more) are commercially available in the *Sketch Engine* (Kilgarriff, 2003-2019; Kilgarriff *et al.*, 2004; Kilgarriff *et al.*, 2014) — currently the leading CQP. Not only is the lexicographer more confident as a result of the corpus revolution in lexicography (Rundell & Stock, 1992a; b; c; Hanks, 2012), users are in turn increasingly being presented with better descriptions of *actual* language, and that presentation works best in digital dictionaries.

RC and BS: From a lexicographic point of view, what are the advantages of the digital format in comparison to traditional print dictionaries?

G-MdS: The digital format offers far more options to 'solve' problematic lexicographical issues, especially for languages with complex morphology. English is trivial lexicographically speaking (it has typically two forms for nouns, four forms for verbs, and very little morphology and thus lemmatisation issues elsewhere); this is very different for many other language families, including for instance the Bantu

languages, which are agglutinative, and where digital dictionaries may truly simplify look-up for both decoding as well as encoding purposes (Prinsloo, 2005; Prinsloo *et al.*, 2012; Prinsloo *et al.*, 2017). Moreover, for polysynthetic languages, such as several of the Amerindian languages, the only truly successful way to lemmatise lexis in a user-friendly way is in a digital product (Frawley *et al.*, 2002).

There are arguably numerous advantages of digital over paper dictionaries, together with numerous digital *Lexicographers' Dreams* (many of which still need to be implemented), but these have been extensively covered in an earlier publication (De Schryver, 2003), to which the reader is referred.

RC and BS: With the invention of computers and the ways in which computers function in the storage and retrieval of information, what was the most important impact this technology has made to the lexicographic practice?

G-MdS: As the late Laurence Urdang correctly observed in an e-mail to me, after having read *Lexicographers' Dreams* (De Schryver, 2003):

> It is easy to speculate or, if you prefer, to dream about what dictionaries might be like in a computer age, but it must be borne in mind that in the early 1960s the world was not yet in the computer age, there were no storage media but punched cards and punched paper tape (and few devices available for converting the latter into magnetic tape, the universal medium for processing and storage), and no personal computers. (personal communication, 18 June 2003)

Beginning in 1959 Urdang saw to the skeletonisation of the thrice-yearly updates of the *American College Dictionary* (DSNA, 2003, p. 6) and the subsequent keyboarding onto paper tape of the *Random House Unabridged* (Stein & Urdang, 1966), "the first [dictionary] to use computers extensively in its editorial preparation" (Urdang, 1984, p. 155). See also Urdang (1966) and Abate (2006, p. 136).

I quote and illustrate this to point out the mind-boggling changes which lexicography (and the computer-world at large) underwent in just half a century: From pre-personal computers with data punched into cards and paper tape, to today's software in the cloud which independently prepares draft corpus-driven dictionary entries for lexicographers to review (Kilgarriff & Rychlý, 2010; Baisa *et al.*, 2019), and/or for users to contribute to via crowdsourcing (Lew, 2014; Rundell, 2017; Kosem, 2019).

Which computational innovation of the past half a century has had the most impact on the lexicographic practice? One actually notices a continuous improvement of all the things a computer can do: more memory, higher speeds, better performance, extra colours, additional multimedia, … and at the same time: ever smaller components and a gradual move from huge mainframe computers maintained by teams to an increasingly more personal and semi-automated experience. All of these levels have been beneficial to dictionary makers: dictionaries may now be compiled faster, and given the near-endless possibilities to check the integrity of the data, they are also ever-more accurate.

Probably the most eventful game changer was when (relational) databases were put into the hands of lexicographers in the 1980s-1990s: handled correctly, dead cross-references for instance, became a ghost of the past. In the mid-2000s, Dictionary Writing Systems started to appear with fully customisable and built-in Document Type Definition (DTD) editors, allowing lexicographers to tailor the dictionary grammar of any project themselves (Joffe & De Schryver, 2005). When the dictionary

**Figure 2.** R.K. Barnhart handling the Remington Rand UNIVAC punch card sorter

[Re-pictured by G-MdS while visiting the 'Barnhart Dictionary Archive' (Lilly Library, Indiana University, Blooming-ton, IN, USA), during DSNA 2019. The text of the caption below is based on the information provided there.]

Lexicographers were among the earliest adopters of computer technology. Of the first digital computer, the UNIVAC (Universal Automatic Computer), a mere 46 were sold, one of them to the Barnharts. In today's prices, they paid a whopping 10 million USD for their system. While working for Clarence L. Barnhart (1900-1993), Laurence Urdang (1927-2008) became the first to make extensive use of the new technology. In this undated photo we see Robert K. Barnhart (1933-2007) visibly satisfied with their Remington Rand UNIVAC punch card sorter, which streamlined the copying and sorting of citations in the Barnhart quotation file.

structure (the DTD or a schema) is kept separate from both the unique article-specific data on the one hand and the repetitive metadata on the other, truly instantaneous tailoring is effectively made possible for digital dictionaries (De Schryver & Joffe, 2005).

RC and BS: In *The Oxford Guide to Practical Lexicography*, the authors B. T. Sue Atkins and Michael Rundell state that: "There is an enormous body of linguistic theory with the potential to help lexicographers to do their jobs more effectively and with greater confidence" (2008, p. 4). As you are well aware, there is no single consensus on the use of linguistic theories to support the lexicographic

Figure 3. Punch card for 'scientific Communism' from the Barnhart quotation file
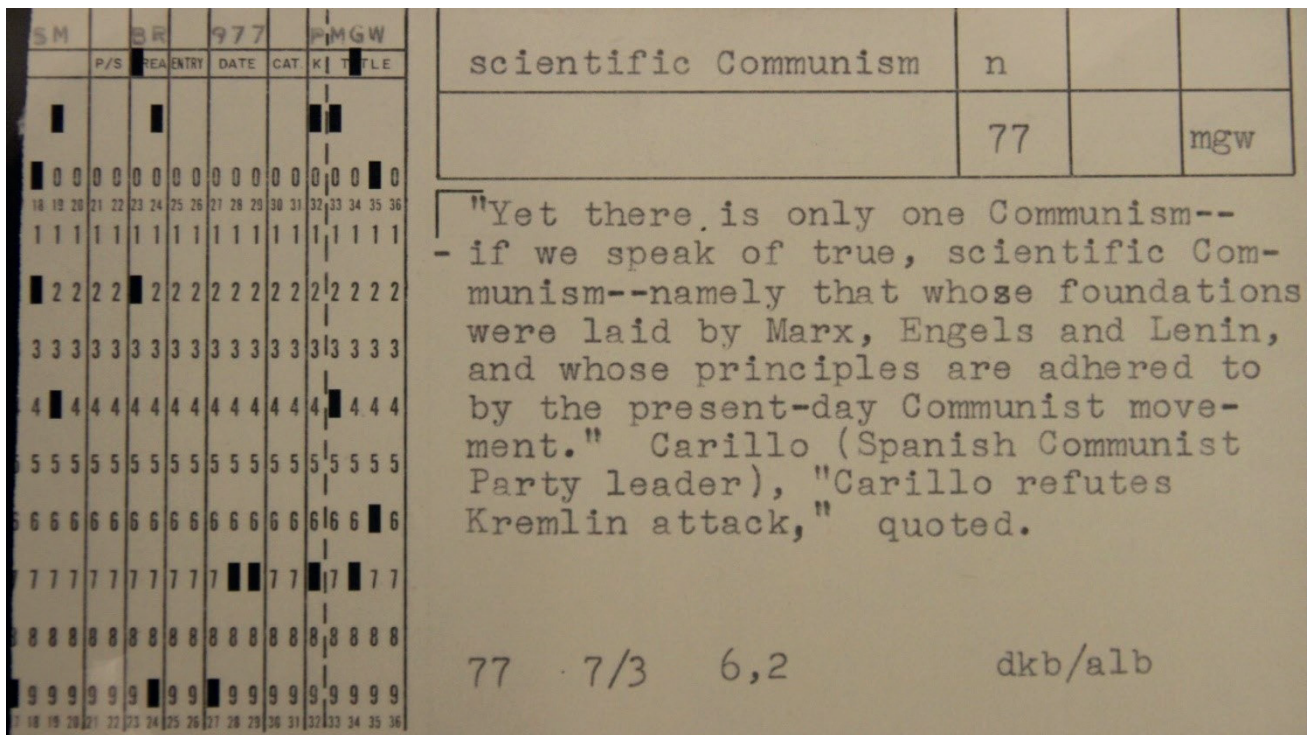
[Photo taken by G-MdS during a visit to the 'Barnhart Dictionary Archive' (Lilly Library, Indiana University, Bloomington, IN, USA), during DSNA 2019.]

Example of a punch card used in lexicography, from the Barnhart quotation file. David K. Barnhart (1995, p. 73) explains: "The head word and some accompanying data were key-punched on the left side of the card; the right side of the card was reserved for the typing of the quotation." A Google search reveals the source to be an article from the New Times, June 24, 1977. The 'dkb' at the bottom-right stands for David K. Barnhart (1941-).

practice. Based on your research and experience, what is your perspective and advice on how lexicographers can more effectively and confidently work?

G-MdS: First: The better dictionaries are indeed compiled using a well-defined and well-applied linguistic theory. Some lexicographic teams cherry-pick different notions from a number of different linguistic frameworks, but this is tricky. Dictionary teams who have not given thought to the linguistic aspects of their work either deliver sloppy lexicographic products, or are natural talents (who have internalised the necessary linguistics).

Second: Your linguistic framework will indeed influence *what* you present and *how* you present that. From the end-user perspective, the result will also *look* and *feel* very different. Take for instance any entry from the *Pattern Dictionary of English Verbs* (Hanks *et al.*, 2019), which uses the technique of Corpus Pattern Analysis (Hanks, 2004), itself based on the linguistic Theory of Norms and Exploitations (Hanks, 2013), and compare this to the corresponding entry/-ies from *FrameNet* (Fillmore *et al.*,

2019), a lexical database of English which is based on a theory of meaning called Frame Semantics (Fillmore, 1976; 1982; Fillmore & Baker, 2009). Same language, both corpus-driven, very different products! They are not necessarily contradictory; Patrick Hanks even wishes to see them as 'complementary' (Hanks *et al.*, 2018, p. 97).

Third: In the previous example, linguistics was needed to provide a theory of meaning; but linguistics is also needed on a far more basic level. Take the parts of speech of a language. The ones in use for English were handed down to us by Latin grammarians, but do they fit English? Not really, but they continue to be used. It becomes even worse when moving to other language families; say Bantu. Apart from the introduction of Bantu-specific parts of speech (POSs), like ideophones (Doke, 1935, p. 119), some lexicographers have suggested to overhaul the word-class system entirely: In his PhD, the lexicographer A.C. Nkabinde (1975) proposes that Zulu, a Bantu language spoken in South Africa, only needs the labels *noun*, *verb*, *ideophone* and *interjection* (with everything else derived from those four POSs).

It should thus be clear that while there is indeed a need for linguistics when compiling dictionaries, as it affects how you study and describe language, there is no need to rank the linguistic theories, and/or to use 'the best' one. The one you use should simply be the most appropriate for the task at hand. As to my advice to lexicographers to be able to work effectively and confidently, I would like to quote a book title from Sinclair (2004): *Trust the Text*. The worst one can do when compiling dictionaries is to use introspection only. You need data (as found in corpora), and a linguistic theory to analyse and synthesise that data.

RC and BS: Some believe that existing e-dictionaries are merely printed dictionaries made available in an electronic platform. Authors, like Wiegand (1998), for example, assume that the task of compiling e-dictionaries requires new theories rather than those used for print dictionaries. Fuertes-Olivera & Bergenholtz (2013), in turn, assert that a theory (or theories) of lexicography should not be limited by the medium (e.g. paper or digital). What is your position?

G-MdS: No doubt, some dictionaries are merely the digital version of what is printed; or rather (as we are going to assume that all future dictionaries are born digitally): the dictionary database has been conceived to produce a certain number of print dictionaries, and these are then offered in digital forms as well, without any substantial differences between the two products (apart from the medium, with some extra bells and whistles in the digital version). Owing to the limits of paper, and after hundreds of years of 'tradition', paper dictionaries tend to look a certain way, and will always look that way — precisely because tradition begets respect. And a dictionary, after all, is a respectable product. This doesn't mean that there are not enormous differences between various paper dictionaries — semasiological vs. onomasiological, mono- vs. bi- vs. multi-lingual, native vs. learners', lexical vs. encyclopaedic, general vs. dialect vs. slang, historical vs. modern, text vs. picture, stem. vs. word vs. phrase vs. idiom vs. proverb etc., word vs. term, frequent vs. common vs. hard, synonym vs. antonym, false-friend and confusable, etymological, rhyme, Scrabble, reverse, frequency, … and any and all combinations of dictionaries along those parameters — the point is that users have come to expect dictionaries to present language data in a certain way, and they want to see this confirmed. Designers of all current Dictionary Writing Systems (DWSs) indeed seem to have 'traditional' paper dictionaries in mind when it comes to outputting data to paper format. However, when it comes to outputting data from a DWS to

any of the many digital formats (HTML, XML, TMX, ODBC, …) one would be mistaken to view the resultant digital dictionary as necessarily and merely being 'a copy' of the paper version.

The first obvious difference is that most modern dictionary databases have both users and machines in mind. It is an open secret that 'our painstakingly compiled dictionary data' is being gobbled up by the likes of Amazon (1994-2019), Google (1997-2019), Facebook (2004-2019) and so on, to feed their sales forces, magic search and translation engines, as well as their advertising efforts. If we do not soon want to become irrelevant as lexicographers, it is high time we work seriously with those Big Data companies.

The second obvious difference is that digital dictionaries have begun to truly look different. The bells and whistles (say audio pronunciations rather than cumbersome IPA symbols or respellings) are now commonly joined by material which may simply not be printed anymore. Take the online *Wordnik* dictionary (McKean *et al.*, 2009-2019), where a search for **whistle** lists so much data and so many cross-references to more data that one can simply not conceive the printed version of it: multiple definitions and etymologies aggregated from other (print) dictionaries, and countless examples automatically grabbed from all over the Internet; hundreds of related forms (362 synonyms, 19 hypernyms, 39 similar contexts, 3 morphological forms, 23 cross-references, 11 rhymes, 189 items from the reverse dictionary, …); many dozens of 'lists' (including 'Bells and Whistles'); a discussion section with user comments and live tweets; the to-be-expected visuals and sounds; and finally some statistics, including its Scrabble score. Could this dictionary be printed? Probably. Will it? No. It won't, because it was born digital (even though it incorporates printed material), and because it changes every single day.

The third obvious difference is that digital dictionaries have started to become (and some already are) fully invisible to their users. Of course authors have always used reference books when they wrote, and in word processors such tools (dictionaries and synonym lists, spelling and grammar corrections, improvements for misused/unlikely forms) merely became available at a click's notice. In the latest augmented writing assistants, however, embedded and invisible digital dictionaries now even raise ethical issues, as they can truly guard what you write and physically stop you from using certain (offensive, racist, non-politically-correct, sexist, …) words (Simonsen, 2019).

Fourthly, digital dictionaries are being used differently. This is most obvious for machine dictionaries. A Google Image Search for **table** will for instance not only return images of various types of table as furniture, but will also offer you images of the periodic table of elements, league table results, casual and formal table settings, table tennis … (where 'table' is indeed part of the combination), but also of items like **snooker** (where 'table' isn't). A Google Video Search for **table** moreover brings up not furniture first, but kids' videos of multiplication tables. This is followed by do-it-yourself instruction videos on how to build various tables as furniture. Clearly, the (assumed) needs of the users are being put front and centre. Also, on mobile devices, the spelling dictionaries of old may for instance be found in the technology known as predictive text. Indeed, our digital natives are so 'locked-in' their mobile devices that the only encounters they ever have with dictionaries on their small screens is as tiny popups that provide more in-context information (definitions, translations, cross-references, …).

Lastly, the ultimate proof that digital dictionaries may have little to do with their paper predecessors is in applications such as picture-to-audio translation dictionaries. All one needs to do, say in a restaurant in Huangcheng (Shanxi, China), is to point one's mobile device at the characters one cannot read on the

menu, to hear what they mean in any other language, or to point one's mobile device at a road sign, say in Annamalai Nagar (Tamil Nadu, India), to make sense of where one is going. In case one wonders: No, these are not random examples, and yes, the user-friendliness of this type of digital lexicography is extreme, as one does not even have to be able to read the script, saving the user from the need to understand the lemmatisation policy used by dictionary makers. This type of digital lexicography also overcomes the age-old problem whereby traditional (paper) dictionaries provide general context-free information, when the user is actually only interested in context-sensitive answers (Varantola, 2002, p. 33).

The fact that digital dictionaries are indeed quite different from paper dictionaries now begs the question: Are they governed by different theories of lexicography? The answer is clearly no! It is no, because I do not believe in a theory of lexicography. There is none, there is no need for one, and there can be no such beast (De Schryver, 2012b, p. 494-498). Inasmuch as people crave theories, and inasmuch as all researchers in lexicography are 'forced' by their funders to work within a theoretical framework, it was 'nice' to have scholars like H.E. Wiegand give it their best, as well as a team in Denmark, but it should be clear to anyone studying their output that merely thinking and theorizing about lexicography, hence pursuing metalexicography (and coining a zillion lexicographic terms in the process, as well as imagining countless usage situations), does not automagically result in a theory of lexicography. It was nice while it lasted, but the interested reader is referred to for instance Piotrowski (2009), Béjoint (2010, p. 381), Kilgarriff (2012), Rundell (2012) and Adamska-Sałaciak (2019) for a sobering view on these futile unfunctional attempts. As nicely put in the title of another classic of our field, lexicography is first and foremost an *Art & Craft* (Landau, 2001).

RC and BS: EURALEX, the European Association for Lexicography, has worldwide reach and membership. As President, you have a unique and privileged perspective about significant projects of lexicography around the globe. Which topics characterize significant projects in the world of lexicography? And what can you identify as essential aspects of these projects?

G-MdS: Allow me to first look back. To my mind, one of the best digital dictionaries for English remains the *Collins COBUILD III on CD-ROM*, which dates from 2001, and is thus already two decades old. This digital dictionary actually brings together five components: a dictionary, a thesaurus, a usage guide, a grammar, and a 5-million-word corpus called Wordbank — each of these components is seamlessly integrated with the others, every word clickable, all of it in a no-frills design. Every version after this third edition, however, failed to include these five components, instead containing a selection of two or three of them, but never again with the Wordbank. Thus, while every second academic study on digital dictionaries talks about the need for or predicts ever-more integration with various other language components and tools, here is a case of backward evolution. It is and remains incomprehensible.

A second aspect that is surprisingly not being picked up by any of the major English dictionary publishers is 'dynamic metalanguage customisation' for multilingual digital dictionaries (De Schryver & Joffe, 2005), whereby the language of such aspects as parts of speech, cross-reference texts, as well as *all* other recurrent labels is automatically adapted to the language of the interface/user. Rather, the online bilingual dictionaries offered by *Oxford*, *Longman*, *Collins* and *Cambridge* simply keep presenting most of these metalanguage strings in the language of the source language only. Again, incomprehensible.

670

Looking at the present now, the very best 'lexicographic resource' may actually very well be the search engine Google (1997-2019). In truth, it is hard to compete with the breadth and speed with which it answers language queries. This only leaves us with the depth.

Human lexicographers are still outcompeting the Big Data companies when it comes to the extreme in-depth detail of their large dictionaries. The subscription-based online OED (2000-2019) is a case in point. For twenty years now, the OED dictionary team has managed to update their data four times a year. While initially a paper dictionary (the last, second edition, was published in 1989), and while the current contents could theoretically still be printed, this is highly unlikely: The OED will live on as a digital dictionary, and the extra features such as graphs (Entry profile, various Timelines, …) and the additional browse paths into the data (such as the Historical Thesaurus index) make it increasingly cumbersome to print all of that out. But, to date, the look and feel of the online OED very much remains that of a paper dictionary in digital format.

Large, detailed amounts of data vs. relatively small mobile screens; how to go about it? Probably the most interesting solution to this conundrum was recently suggested by a team of researchers from the *Institut für Deutsche Sprache* at eLex 2019. According to Michaelis *et al.* (2019, p. 43): "variations and rearrangements of the basic building blocks of a dictionary entry can be used to create new and completely different interfaces". The idea is basically to cut up huge dictionary articles into tiles going both from top to bottom (at which point one best scrolls a mobile device vertically) and from left to right (at which point one best scrolls a mobile device horizontally). It is thus suggested that the resultant two-dimensional space of tiles (one tile per sense, and one tile each for aspects such as synonyms, etymology, grammar, audio, combinations, …) is 'navigated'. In order not to lose track, a 'miniature map' of the original article may be presented on the side, with a pointer indicating where in the large article/space one is navigating.

This is clearly an original idea, and its efficacy should be tested empirically. However (and as I also suggested to the two co-authors present at eLex 2019, Sascha Wolfer and Carolin Müller-Spitzer) even such a tile layout still assumes that people look up in digital dictionaries in ways similar to a paper-dictionary experience. Rather, one should bring this novel type of presentation to its logical conclusion, and admit that modern dictionary users do not 'look up' anymore but 'search' (De Schryver, 2012a, p. 130). Hence, rather than excepting a dictionary user to start browsing through the tiles, a user's *search query* should immediately lead to that tile which contains the answer — much like the 'I'm Feeling Lucky' button of Google, but with this difference that the neighbouring tiles are also easily accessible (at a swipe), just in case one felt wrongly lucky.

Today's most interesting lexicographic projects, then, are those where one truly thinks out of the box, in order to come up with novel ways to both present and get at dictionary data. One such group of original thinkers is undoubtedly the team at the *Institut für Deutsche Sprache* in Mannheim (Germany). Essential, too, is to test one's proposals, for which the team in Mannheim also has an excellent track record (see e.g. the edited collections by Müller-Spitzer, 2014; Klosa & Müller-Spitzer, 2016). Another team which does top-notch research in this regard, are the lexicographers at the *Adam Mickiewicz University* in Poznań (Poland), especially Robert Lew (e.g. Lew, 2004; Lew & Dziemianko, 2006a; b; Lew & Doroszewska, 2009; Lew, 2010; 2011; 2012; 2013; Lew *et al.*, 2013; Lew &

Figure 4. S. Wolfer muses on lexicographic data spread out in two dimensions
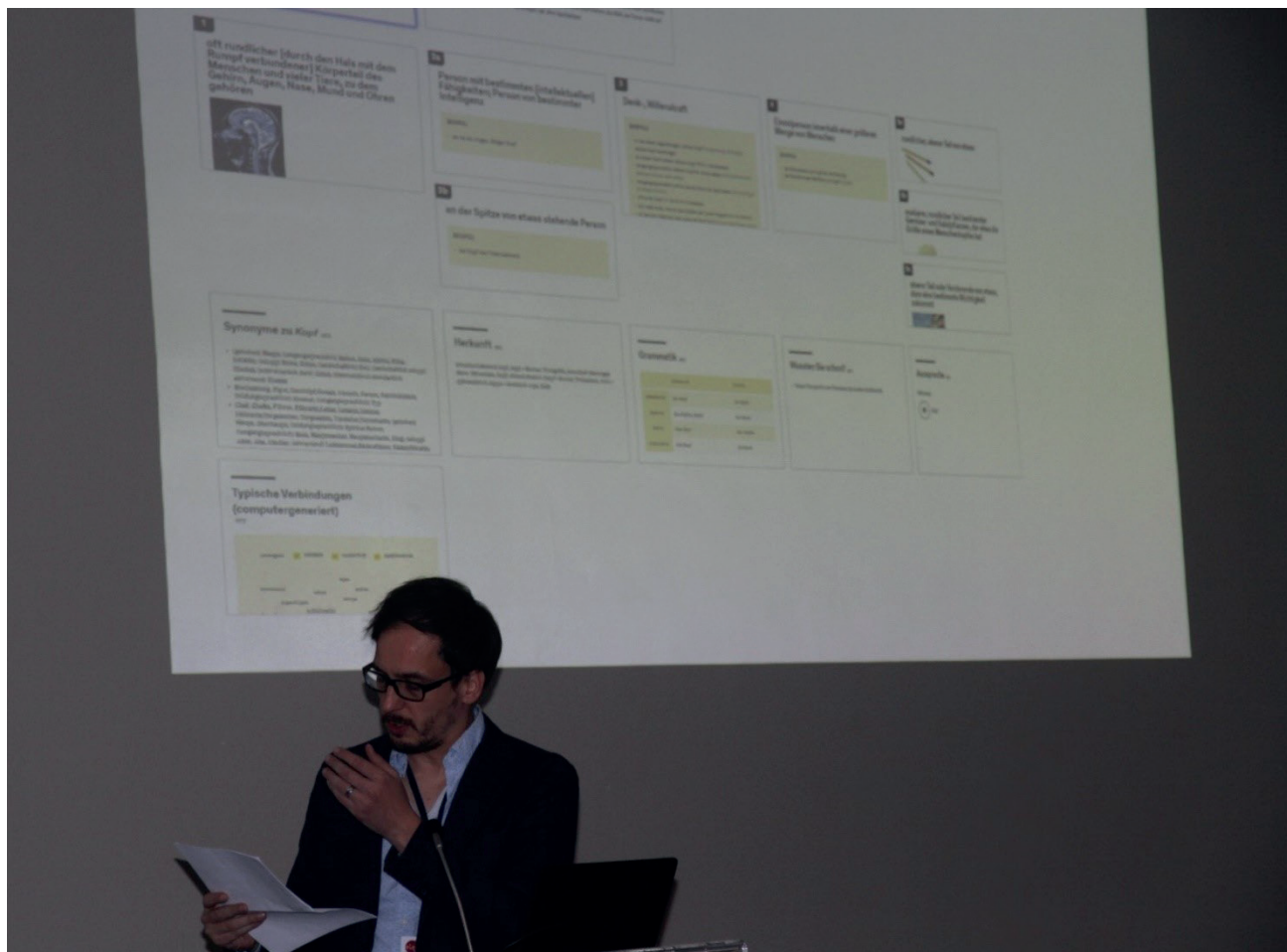
[Photo taken by G-MdS at eLex 2019. A video of the presentation is available from: https://www.youtube.com/watch?v=rjirfjTbQo4, 1:29:20 > 2:00:31]

*Sascha Wolfer swipes through one of the 'Sintra variations', being a way to navigate long dictionary articles on mobile devices in a two-dimensional space of tiles; here illustrated for the German article **der Kopf** 'the head'.*

De Schryver, 2014; Lew *et al.*, 2018) and Anna Dziemianko (e.g. Dziemianko, 2006; 2010; 2011a; b; 2012a; b; 2014; 2015; 2019a; b; c).

RC and BS: The 19th EURALEX Conference is set to take place from September 8-12, 2020 in Alexandroupolis, Greece. The conference motto is: "Lexicography for Inclusion." What it the essential purpose of this motto? In consideration of this topic of inclusion and its meaning, to what extent has progress been made in digital lexicography?

G-MdS: Given this question, it is important to point out that one should not confuse eLex with EURALEX. The first is a series of conferences facilitated by a handful of scholars in Ljubljana (Slovenia) and

Brno (Czech Republic), the second is an Association with a Constitution, a Board, paid-up Members, a Journal (the *International Journal of Lexicography*), etc. which *inter alia* holds conferences. Both eLex and EURALEX produce Conference Proceedings. The aim of the eLex conferences is to bring researchers on digital lexicography together, for all of the world's languages; in practice the eLex conferences indeed manage to focus on digital lexicography, but mostly deal with European languages. The aim of the EURALEX conferences is to welcome all types of lexicography, for the European languages; in practice the EURALEX conferences increasingly cover a lot of digital lexicography, and this for all of the world's languages.

As to the moto of Euralex 2020, it is indeed 'Lexicography for Inclusion'. The intention is to cover all sorts of topics and types of dictionary-making which have not featured often enough at the EURALEX conferences to date, including but not limited to dictionaries for the visually-impaired, translation dictionaries for migrants, idioticons for non-standard varieties, code for contemporary queer cultures, etc. This motto is thus not necessarily linked to digital lexicography per se. That said, we are confident that Inclusion Dictionaries have moved digital as well, and we expect to learn about that in Alexandroupolis.

RC and BS: Considering your experience and position of influence in the world of lexicography, what are your thoughts about the future of digital lexicography?

G-MdS: If lexicography is going to survive at all; *digital* lexicography is the only type of lexicography that will survive. Our grandchildren will not often hold or touch paper anymore, if any. Everything will be retro-digitized. Books and anything on paper will be for nerds and/or researchers who will have the task to comb dusty archives.

As to the future of digital lexicography, the current trend will, as I argued elsewhere: "result in reference works beyond all recognition" (De Schryver, 2003, p. 143). This has been echoed by many other scholars since, for instance:

> While it is fairly uncontroversial that people will continue to have lexical needs in natural communication as well as in more or less artificial learning contexts, it is much less certain that dictionaries will persist for much longer, at least in the form we know them today. Rather, it seems likely that dictionaries will increasingly become absorbed into more general digital tools designed to provide assistance with communication, expression, and information searching. (Lew 2015, p. 7)

At the same time, and this may be an aspect that is not often enough considered, future lexicographers may very well have to teach linguists something, as argued in the Position Paper by Patrick Hanks at EURALEX 2016:

> Noam Chomsky (surely the greatest living polemicist) famously dismissed lexicography as 'mere butterfly collecting', and one can see how he could have got this impression. In 1972 he dismissed Saussurean structuralism as an "impoverished and thoroughly inadequate conception of language". Against this, anyone interested in meaning might be forgiven for dismissing Chomskyan linguistics, which has so little that is useful or accurate to say about meaning, as an even more impoverished and inadequate conception of language.
>
> A result of such polemics is that, in many universities, courses in linguistics do not pay adequate attention to words, phraseology, and meaning. Lexicography can help to change this

situation, but it will take time, perseverance, and, I suggest, a continuing change of focus. (Hanks, 2016, p. 2)

RC and BS: If you believe that there are other essential issues to address, could you kindly use this space to posit a question?

G-MdS: In *Lexicographers' Dreams* (De Schryver, 2003) and its follow-up *Intelligent Lexicography* (De Schryver, 2010b, p. 584) I proposed a new type of dictionary: "one in which the potential is explored to link an automatically derived dynamic user profile to the proffered multimedia lexicographic output". Given that such adaptive and intelligent dictionaries would use artificial intelligence components, and given that AI is the hype of the day, how much longer will it take before a smart dictionary team will take on the challenge?

References

ABATE, F. 2006. Alphabetic order and disorder. *Dictionaries: Journal of The Dictionary Society of North America*, **27**:133-136. https://doi.org/10.1353/dic.2006.0008

ADAMSKA-SAŁACIAK, A. 2019. Lexicography and theory: Clearing the ground. *International Journal of Lexicography*, **32**(1):1-19. https://doi.org/10.1093/ijl/ecy017

AMAZON. 1994-2019. *Amazon – E-commerce, cloud computing, digital streaming and artificial intelligence company*. Available online at: https://www.amazon.com/.

ATKINS, B.T.S. (ed.). 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators* (Lexicographica. Series Maior 88). Tübingen, Max Niemeyer, 214 p. https://doi.org/10.1515/9783110929997

ATKINS, B.; RUNDELL, M. 2008. *The Oxford Guide to Practical Lexicography*. 1ed. New York, Oxford University Press, 540 p.

BAISA, V.; BLAHUŠ; M.; CUKR, M.; HERMAN, O.; JAKUBÍČEK, M.; KOVÁŘ, V.; MEDVEĎ, M.; MĚCHURA, M.; RYCHLÝ, P.; SUCHOMEL, V. 2019. Automating dictionary production: A Tagalog-English-Korean dictionary from scratch. In: Kosem, Iztok, Tanara Zingano Kuhn, Margarita Correia, José Pedro Ferreira, Maarten Jansen, Isabel Pereira, Jelena Kallas, Miloš Jakubíček, Simon Krek & Carole Tiberius (eds). *Electronic Lexicography in the 21st Century (eLex 2019): Smart lexicography. Conference proceedings. Sintra, Portugal, 1-3 October 2019*. Brno, Lexical Computing, p. 805-818.

BARNHART, D.K. 1995. Clarence L. Barnhart and quotations for new words in American dictionaries. *Dictionaries: Journal of The Dictionary Society of North America*, **16**:65-78. https://doi.org/10.1353/dic.1995.0002

BÉJOINT, H. 2010. *The Lexicography of English. From Origins to Present*. New York, Oxford University Press, 458 p.

674

BOGAARDS, P. 1998. Scanning long entries in learner's dictionaries. In: Fontenelle, Thierry, Philippe Hiligsmann, Archibald Michiels, André Moulin & Siegfried Theissen (eds). *ACTES EURALEX'98 PROCEEDINGS, Communications soumises à EURALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique / Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*. Liège, English and Dutch Departments, University of Liège, p. 555-563.

BOSTOEN, K.; DE SCHRYVER, G-M. 2015. Linguistic innovation, political centralization and economic integration in the Kongo kingdom: Reconstructing the spread of prefix reduction. *Diachronica*, **32**(2):139-185 + 13 pages of supplementary material online. https://doi.org/10.1075/dia.32.2.01bos

BOSTOEN, K.; DE SCHRYVER, G-M. 2018. Seventeenth-century Kikongo is not the ancestor of present-day Kikongo. In: Bostoen, Koen & Inge Brinkman (eds). *The Kongo Kingdom: The Origins, Dynamics and Cosmopolitan Culture of an African Polity*. Cambridge, Cambridge University Press, p. 60-102. https://doi.org/10.1017/9781108564823

BRUSCIOTTO, G. 1659. *Regulae quaedam pro difficillimi Congensium idiomatis faciliori captu ad grammaticae normam redactae. A F. Hyacintho Brusciotto a Vetralla Concionatore Capuccino Regni Congi Apostolicae Missionis Praefecto [Some rules for the more easy understanding of the most difficult idiom of the people of the Congo, brought into the form of a grammar]*. Rome, Sacra Congregatio de Propaganda Fide, 106 p.

CARDOSO, M. 1624. *Dovtrina Christaã. Composta pelo P. Marcos Iorge da Companhia de IESV Doutor em Theologia. Acrescentada pelo Padre Ignacio Martinz da mesma Companhia Doutor Theologo. De nouo traduzida na lingoa do Reyno de Congo, por ordem do P. Mattheus Cardoso Theologo, da Companhia de IESV, natural da cidade de Lisboa. Ao muito poderoso, & catholico Rey de Congo dom Pedro Affonso Segundo deste nome [Christian Doctrine. Composed by Father Marcos Jorge of the Society of Jesus, PhD. Augmented by Father Ignacio Martinz of the same Society, PhD. Translated into the language of the Kongo Kingdom under the supervision of Mattheus Cardoso, Theologian of the Society of Jesus, born in the city of Lisbon. To the very powerful and catholic King of the Kongo Dom Pedro Affonso II of this name]*. Lisbon, Geraldo da Vinha, 278 p.

DE PAUW, G.; DE SCHRYVER, G-M; WAGACHA, P.W. 2006-2019. *AfLaT – African Language Technology*. Available online at: http://aflat.org/.

DE SCHRYVER, G-M. 2003. Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography*, **16**(2):143-199. https://doi.org/10.1093/ijl/16.2.143

DE SCHRYVER, G-M. 2010a. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford*. Cape Town, Oxford University Press Southern Africa, xii + 596 p.

DE SCHRYVER, G-M. 2010b. State-of-the-art software to support Intelligent Lexicography. In: 朱瑞平 [Zhu, Ruiping] (ed.). 中華字典研究 *[Chinese Lexicographic Research]*, **2**:584-599. [s.l.], 中國社科 [Chinese Social Sciences].

DE SCHRYVER, G-M. 2012a. Lexicography in the crystal ball: Facts, trends and outlook. In: Fjeld, Ruth V. & Julie M. Torjusen (eds). *Proceedings of the 15th EURALEX International Congress, 7-11*

August, 2012, Oslo. Oslo, Department of Linguistics and Scandinavian Studies, University of Oslo, p. 93-163.

DE SCHRYVER, G-M. 2012b. Trends in twenty-five years of academic lexicography. *International Journal of Lexicography*, **25**(4):464-506 + 42 pages of supplementary material online. https://doi.org/10.1093/ijl/ecs030

DE SCHRYVER, G-M; JOFFE, D. 2004. On how electronic dictionaries are really used. In: Williams, Geoffrey & Sandra Vessier (eds). *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient, Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, p. 187-196.

DE SCHRYVER, G-M; JOFFE, D. 2005. Dynamic metalanguage customisation with the dictionary application TshwaneLex. In: Kiefer, Ferenc, Gábor Kiss & Júlia Pajzs (eds). *Papers in Computational Lexicography, COMPLEX 2005*. Budapest, Linguistics Institute, Hungarian Academy of Sciences, p. 190-199.

DE SCHRYVER, G-M; PRINSLOO, D.J. 2000. The compilation of electronic corpora, with special reference to the African languages. *Southern African Linguistics and Applied Language Studies*, **18**(1-4):89-106.

DE SCHRYVER, G-M; REYNOLDS, M. 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English / Oxford isiXhosa-isiNgesi English-isiXhosa Isichazi-magama Sesikolo*. Cape Town, Oxford University Press Southern Africa, lxx + 562 p.

DE SCHRYVER, G-M; WOLFER, S.; LEW, R. 2019. The relationship between dictionary look-up frequency and corpus frequency revisited: A log-file analysis of a decade of user interaction with a Swahili-English dictionary. *GEMA Online® Journal of Language Studies*, **19**(4):1-27 + supplementary material online.

DESCOURVIÈRES, J.J. 1772. *Dictionnaire françois et congo*. London, British Museum, manuscript n° 33779, main copyist Belgarde.

DESCOURVIÈRES, J.J. 1773. *Dictionnaire françois et congo*. Besançon, Bibliothèque Municipale, manuscript n° 525, main copyist Cuénot.

DESCOURVIÈRES, J.J. 1775. *Dictionnaire congo et françois*. Besançon, Bibliothèque Municipale, manuscript n° 524, main copyist Cuénot.

DESCOURVIÈRES, J.J. 1776. *Essai d'une grammaire congo, suivant l'accent de Kakongo*. Besançon, Bibliothèque Municipale, manuscript n° 523, copyist Cuénot.

DOKE, C.M. 1935. *Bantu Linguistic Terminology*. London, Longmans, Green & Co, 237 p.

DSNA. 2003. *DSNA Newsletter 27(2)*. Madison, WI, University of Wisconsin-Madison, 8 p.

DZIEMIANKO, A. 2006. *User-friendliness of verb syntax in pedagogical dictionaries of English*. (Lexicographica. Series Maior 130). Tübingen, Max Niemeyer, 229 p. https://doi.org/10.1515/9783110941098

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. https://doi.org/10.1093/ijl/ecp040

DZIEMIANKO, A. 2011a. Does dictionary form really matter? In: Akasu, Kaoru & Satoru Uchida (eds). *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference. Kyoto Terrsa, Kyoto, Japan, August 22-24, 2011*. Kyoto, The Asian Association for Lexicography, p. 92-101.

DZIEMIANKO, A. 2011b. User-friendliness of noun and verb coding systems in pedagogical dictionaries of English: A case of Polish learners. *International Journal of Lexicography*, **24**(1):50-78. https://doi.org/10.1093/ijl/ecq037

DZIEMIANKO, A. 2012a. On the use(fulness) of paper and electronic dictionaries. In: Granger, Sylviane & Magali Paquot (eds). *Electronic Lexicography*. Oxford, Oxford University Press, p. 319-341. https://doi.org/10.1093/acprof:oso/9780199654864.003.0015

DZIEMIANKO, A. 2012b. Why one and two do not make three: Dictionary form revisited. *Lexikos*, **22**:195-216. https://doi.org/10.5788/22-1-1003

DZIEMIANKO, A. 2014. On the presentation and placement of collocations in monolingual English learners' dictionaries: Insights into encoding and retention. *International Journal of Lexicography*, **27**(3):259-279. https://doi.org/10.1093/ijl/ecu012

DZIEMIANKO, A. 2015. Colours in online dictionaries: A case of functional labels. *International Journal of Lexicography*, **28**(1):27-61. https://doi.org/10.1093/ijl/ecu028

DZIEMIANKO, A. 2019a. Homogeneous or heterogeneous? Insights into signposts in learners' dictionaries. *International Journal of Lexicography*, **32**(4):432-457. https://doi.org/10.1093/ijl/ecz011

DZIEMIANKO, A. 2019b. The role of online dictionary advertisements in language reception, production, and retention. *ReCALL: The Journal of the European Association for Computer Assisted Language Learning*, **31**(1):5-22. https://doi.org/10.1017/S0958344018000149

DZIEMIANKO, A. 2019c. Smart advertising and online dictionary usefulness. In: Kosem, Iztok & Tanara Zingano Kuhn (eds). *Electronic Lexicography in the 21st Century (eLex 2019): Smart lexicography. Book of abstracts. Sintra, Portugal, 1-3 October 2019*. Brno, Lexical Computing, p. 25-26.

FACEBOOK. 2004-2019. *Facebook – Online social media and social networking service*. Available online at: https://www.facebook.com.

FILLMORE, C.J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, **280**(1):20-32. https://doi.org/10.1111/j.1749-6632.1976.tb25467.x

FILLMORE, C.J. 1982. Frame Semantics. In: The Linguistic Society of Korea (ed.). *Linguistics in the Morning Calm*. Seoul, Hanshin Publishing Co., p. 111-137.

FILLMORE, C.J.; BAKER, C.F. 2009. A frames approach to semantic analysis. In: Heine, Bernd & Heiko Narrog (eds). *The Oxford Handbook of Linguistic Analysis*. New York, NY, Oxford University Press, p. 313-341.

FILLMORE, C.J. *et al*. 2019. *The FrameNet project*. Available online at: https://framenet.icsi.berkeley.edu/. FOX, G. 1987. The case for examples. In: Sinclair, John M. (ed.). *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London, Collins ELT, p. 137-149.

FRAWLEY, W.; HILL, K.C.; MUNRO, P. 2002. *Making Dictionaries. Preserving Indigenous Languages of the Americas*. Berkeley, CA, University of California Press, 455 p.

FUERTES-OLIVERA, P.A.; BERGENHOLTZ, H. (eds). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 1ed. London; New York, Continuum, 341 p.

GOOGLE. 1997-2019. *Google – Web search engine*. Available online at: https://google.com/.

HANKS, P. 2002. Mapping meaning onto use. In: Corréard, Marie-Hélène (ed.). *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. [s.l.], Euralex, p. 156-198.

HANKS, P. 2004. Corpus Pattern Analysis. In: Williams, Geoffrey & Sandra Vessier (eds). *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient, Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, p. 87-97.

HANKS, P. 2012. The corpus revolution in lexicography. *International Journal of Lexicography*, **25**(4):398-436. https://doi.org/10.1093/ijl/ecs026

HANKS, P. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA, The MIT Press, 462 p. https://doi.org/10.7551/mitpress/9780262018579.001.0001

HANKS, P. 2016. *Position paper for Tbilisi Round Table*. 3 p. Available online at: https://www.euralex.org/wp-content/uploads/2016/10/Position_Paper.pdf.

HANKS, P. *et al*. 2019. *Pattern Dictionary of English Verbs*. Available online at: http://www.pdev.org.uk/.

HANKS, P.; EL MAAROUF, I.; OAKES, M. 2018. Flexibility of multiword expressions and Corpus Pattern Analysis. In: Sailer, Manfred & Stella Markantonatou (eds). *Multiword Expressions: Insights from a Multilingual Perspective*. Berlin, Language Science Press, p. 93-119.

HARTMANN, R.R.K. 1987. Four perspectives on dictionary use: A critical review of research methods. In: Cowie, Anthony P. (ed.). *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985* (Lexicographica. Series Maior 17). Tübingen, Max Niemeyer, p. 246-256.

HURSKAINEN, A. 1992. A two-level computer formalism for the analysis of Bantu morphology: An application to Swahili. *Nordic Journal of African Studies*, **1**(1):87-122.

JOFFE, D.; DE SCHRYVER, G-M. 2002-2019. *TLex Suite – Dictionary Compilation Software*. Available online at: http://tshwanedje.com/tshwanelex/.

JOFFE, D.; DE SCHRYVER, G-M. 2005. Representing and describing words flexibly with the dictionary application TshwaneLex. In: Ooi, Vincent B.Y., Anne Pakir, Ismail S. Talib, Lynn Tan, Peter K.W. Tan & Ying Y. Tan (eds). *Words in Asian Cultural Contexts, Proceedings of the 4th Asialex Conference, 1-3 June 2005, M Hotel, Singapore*. Singapore, Department of English Language and Literature & Asia Research Institute, National University of Singapore, p. 108-114.

KANETA, T. 2011. Folded or unfolded: Eye-tracking analysis of L2 learners' reference behavior with different types of dictionary. In: Akasu, Kaoru & Satoru Uchida (eds). *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference. Kyoto Terrsa, Kyoto, Japan, August 22-24, 2011*. Kyoto, The Asian Association for Lexicography, p. 219-224.

KEMMER, K. 2014a. *Illustrationen im Onlinewörterbuch. Text-Bild-Relationen im Wörterbuch und ihre Empirische Untersuchung*. (Amades 47). Mannheim, Institut für Deutsche Sprache, 485 p.

KEMMER, K. 2014b. Rezeption der Illustration, jedoch Vernachlässigung der Paraphrase? Ergebnisse einer Benutzerbefragung und Blickbewegungsstudie. In: Müller-Spitzer, Carolin (ed.). *Using Online Dictionaries* (Lexicographica. Series Maior 145). Berlin, Walter de Gruyter, p. 251-278.

KILGARRIFF, A. 2003-2019. *Sketch Engine*. Available online at: https://www.sketchengine.eu/.

KILGARRIFF, A. 2012. Book Review: P.A. Fuertes-Olivera and H. Bergenholtz, eds. 2011. e-Lexicography: The Internet, Digital Initiatives and Lexicography. *Kernerman Dictionary News*, **20**:26-29.

KILGARRIFF, A.; BAISA, V.; BUŠTA, J.; JAKUBÍČEK, M.; KOVÁŘ, V.; MICHELFEIT, J.; RYCHLÝ, P.; SUCHOMEL, V. 2014. The Sketch Engine: Ten years on. *Lexicography: Journal of ASIALEX*, **1**(1):7-36. https://doi.org/10.1007/s40607-014-0009-9

KILGARRIFF, A.; HUSÁK, M.; MCADAM, K.; RUNDELL, M.; RYCHLÝ, P. 2008. GDEX: Automatically finding good dictionary examples in a corpus. In: Bernal, Elisenda & Janet DeCesaris (eds). *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)* (Sèrie Activitats 20). Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, p. 425-432.

KILGARRIFF, A.; KOVÁŘ, V.; RYCHLÝ, P. 2010. Tickbox Lexicography. *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve, Centre for English Corpus Linguistics, Université catholique de Louvain, p. 411-418.

KILGARRIFF, A.; RYCHLÝ, P. 2010. Semi-automatic dictionary drafting. In: De Schryver, Gilles-Maurice (ed.). *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks* (Linguistics Series). Kampala, Menha Publishers, p. 299-312.

KILGARRIFF, A.; RYCHLÝ, P.; SMRZ, P.; TUGWELL, D. 2004. The Sketch Engine. In: Williams, Geoffrey & Sandra Vessier (eds). *Proceedings of the Eleventh EURALEX International Congress, EU-*

RALEX 2004, Lorient, France, July 6-10, 2004. Lorient, Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, p. 105-115.

KILGARRIFF, A.; TUGWELL, D. 2001. WORD SKETCH: Extraction and display of significant collocations for lexicography. *Workshop Proceedings. COLLOCATION: Computational Extraction, Analysis and Exploitation (39th Annual Meeting of the Association for Computational Linguistics and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, 7 July 2001)*. p. 32-38.

KLOSA, A.; MÜLLER-SPITZER, C. (eds). 2016. *Internetlexikografie. Ein Kompendium*. Berlin, Walter de Gruyter, 347 p. https://doi.org/10.1515/9783050095615

KOSEM, I. 2019. Responsive dictionaries: Bringing dictionaries (and lexicographers) and users closer together. *15 konferensen om lexikografi i Norden, 4−7 juni 2019. Abstrakt*. Helsinki, Institute for the Languages of Finland & Nordic Association for Lexicography, p. 4-5. A video of the presentation is available from: https://www.youtube.com/watch?v=vJGXwvK682E (see esp. 29:39 > 45:36).

LANDAU, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography.* 2 ed. Cambridge, Cambridge University Press, 370 p.

LAUFER, B.; MELAMED, L. 1994. Monolingual, bilingual and 'bilingualised' dictionaries: Which are more effective, for what and for whom? In: Martin, Willy, Willem Meijs, Margreet Moerland, Elsemiek ten Pas, Piet van Sterkenburg & Piek Vossen (eds). *Euralex 1994 Proceedings. Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*. Amsterdam, Vrije Universiteit, p. 565-576.

LEW, R. 2004. *Which Dictionary for Whom? Receptive use of bilingual, monolingual and semi-bilingual dictionaries by Polish learners of English*. Poznań, Motivex, x + 217 p.

LEW, R. 2010. Users take shortcuts: Navigating dictionary entries. In: Dykstra, Anne & Tanneke Schoonheim (eds). *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Leeuwarden, Fryske Akademy, p. 1121-1132.

LEW, R. 2011. Studies in dictionary use: Recent developments. *International Journal of Lexicography*, **24**(1):1-4. https://doi.org/10.1093/ijl/ecq044

LEW, R. 2012. How can we make electronic dictionaries more effective? In: Granger, Sylviane & Magali Paquot (eds). *Electronic lexicography*. Oxford, Oxford University Press, p. 343-361. https://doi.org/10.1093/acprof:oso/9780199654864.003.0016

LEW, R. 2013. Online dictionary skills. In: Kosem, Iztok, Jelena Kallas, Polona Gantar, Simon Krek, Margit Langemets & Maria Tuulik (eds). *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana, Trojina, Institute for Applied Slovene Studies / Tallinn, Eesti Keele Instituut, p. 16-31.

LEW, R. 2014. User-generated content (UGC) in English online dictionaries. *OPAL – Online Publizierte Arbeiten Zur Linguistik*, **4**:8-26.

LEW, R. 2015. Dictionaries and their users. In: Hanks, Patrick & Gilles-Maurice de Schryver (eds). *International Handbook of Modern Lexis and Lexicography*. Berlin, Springer, p. 1-9. https://doi.org/10.1007/978-3-642-45369-4_11-1

LEW, R.; DE SCHRYVER, G-M. 2014. Dictionary users in the digital revolution. *International Journal of Lexicography*, **27**(4):341-359. https://doi.org/10.1093/ijl/ecu011

LEW, R.; DOROSZEWSKA, J. 2009. Electronic dictionary entries with animated pictures: Lookup preferences and word retention. *International Journal of Lexicography*, **22**(3):239-257. https://doi.org/10.1093/ijl/ecp022

LEW, R.; DZIEMIANKO, A. 2006a. A new type of folk-inspired definition in English monolingual learners' dictionaries and its usefulness for conveying syntactic information. *International Journal of Lexicography*, **19**(3):225-242. https://doi.org/10.1093/ijl/ecl011

LEW, R.; DZIEMIANKO, A. 2006b. Non-standard dictionary definitions: What they cannot tell native speakers of Polish. *Cadernos de Tradução*, **18**(2):275-294.

LEW, R.; GRZELAK, M.; LESZKOWICZ, M. 2013. How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. *Lexikos*, **23**:228-254. https://doi.org/10.5788/23-1-1213

LEW, R.; KAŹMIERCZAK, R.; TOMCZAK, E; LESZKOWICZ, M. 2018. Competition of definition and pictorial illustration for dictionary users' attention: An eye-tracking study. *International Journal of Lexicography*, **31**(1):53-77. https://doi.org/10.1093/ijl/ecx002

MCKEAN, E.; BARRETT, G.; MONTOYA, O. 2009-2019. *Wordnik – The world's biggest online English dictionary*. Available online at: https://www.wordnik.com/.

MICHAELIS, F.; MÜLLER-SPITZER, C.; WOLFER, S. 2019. The Sintra variations – Thinking outside the box in designing online dictionaries. In: Kosem, Iztok & Tanara Zingano Kuhn (eds). *Electronic Lexicography in the 21st Century (eLex 2019): Smart lexicography. Book of abstracts. Sintra, Portugal, 1-3 October 2019*. Brno, Lexical Computing, p. 43-44.

MÜLLER-SPITZER, C. (ed.). 2014. *Using Online Dictionaries* (Lexicographica. Series Maior 145). Berlin, Walter de Gruyter, 387 p. https://doi.org/10.1515/9783110341287

MÜLLER-SPITZER, C.; MICHAELIS, F.; KOPLENIG, A. 2014. Evaluation of a new web design for the dictionary portal OWID: An attempt at using eye-tracking technology. In: Müller-Spitzer, Carolin (ed.). *Using Online Dictionaries* (Lexicographica. Series Maior 145). Berlin, Walter de Gruyter, p. 207-228.

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, 156 p. https://doi.org/10.1515/9783110946031

NKABINDE, A.C. 1975. *A Revision of the Word Categories in Zulu*. PhD dissertation. Pretoria, UNISA, 184 p.

OED. 2000-2019. *Oxford English Dictionary – The definitive record of the English language*. Available online at: https://www.oed.com/.

PIOTROWSKI, T. 2009. Book Review: S. Tarp. 2008. Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. *International Journal of Lexicography*, **22**(4):480-486. https://doi.org/10.1093/ijl/ecp030

PRINSLOO, D.J. 1991. Towards computer-assisted word frequency studies in Northern Sotho. *South African Journal of African Languages*, **11**(2):54-60. https://doi.org/10.1080/02572117.1991.10586891

PRINSLOO, D.J. 2005. Electronic dictionaries viewed from South Africa. *Hermes, Journal of Linguistics*, **34**:11-35. https://doi.org/10.7146/hjlcb.v18i34.25798

PRINSLOO, D.J.; BOTHMA, T.J.D.; HEID, U.; PRINSLOO, D.J. 2017. Direct user guidance in e-dictionaries for text production and text reception — The verbal relative in Sepedi as a case study. *Lexikos*, **27**:403-426. https://doi.org/10.5788/27-1-1408

PRINSLOO, D.J.; HEID, U.; BOTHMA, T.J.D.; FAAß, G. 2012. Devices for information presentation in electronic dictionaries. *Lexikos*, **22**:290-320. https://doi.org/10.5788/22-1-1009

RUNDELL, M. 2012. It works in practice but will it work in theory? The uneasy relationship between lexicography and matters theoretical. In: Fjeld, Ruth V. & Julie M. Torjusen (eds). *Proceedings of the 15th EURALEX International Congress, 7-11 August, 2012, Oslo*. Oslo, Department of Linguistics and Scandinavian Studies, University of Oslo, p. 47-92.

RUNDELL, M. 2017. Dictionaries and crowdsourcing, wikis and user-generated content. In: Hanks, Patrick & Gilles-Maurice de Schryver (eds). *International Handbook of Modern Lexis and Lexicography*. Berlin, Springer, p. 1-16. https://doi.org/10.1007/978-3-642-45369-4_26-1

RUNDELL, M.; STOCK, P. 1992a. The corpus revolution 1. The first in a series of three reports on the development and use of electronic language corpora and their impact on dictionaries. *English Today, The international review of the English language*, **8**(2):9-14. https://doi.org/10.1017/S026607840000626X

RUNDELL, M.; STOCK, P. 1992b. The corpus revolution 2. A consideration of the practical benefits to English-language lexicographers of the evidence derived from computer corpora (second article of three). *English Today, The international review of the English language*, **8**(3):21-29. https://doi.org/10.1017/S0266078400006520

RUNDELL, M.; STOCK, P. 1992c. The corpus revolution 3. A consideration of the prospects and potential of corpus-and-concordance lexicography (third article of three). *English Today, The international review of the English language*, **8**(4):45-51. https://doi.org/10.1017/S0266078400006751

SIMONSEN, H.K. 2009a. Se – og du skal finde. En eye tracking-undersøgelse med særlig fokus på de leksikografiske funktioner. *Nordiske studier i leksikografi*, **10**:274-288.

SIMONSEN, H.K. 2009b. *Vertical or horizontal? That is the question: An eye-track study of data presentation in Internet dictionaries*. Paper presented at the Eye-to-IT Conference on Translation Proces-

ses, Sentence Processing and the Bilingual Mental Lexicon, Business School, Copenhagen, Denmark, April 25–29, 2009.

SIMONSEN, H.K. 2011. User consultation behaviour in Internet dictionaries: An eye-tracking study. *Hermes, Journal of Language and Communication Studies*, **46**:75-101. https://doi.org/10.7146/hjlcb. v24i46.97370

SIMONSEN, H.K. 2019. Augmented writing: New opportunities for lexicography? In: Kosem, Iztok & Tanara Zingano Kuhn (eds). *Electronic Lexicography in the 21st Century (eLex 2019): Smart lexicography. Book of abstracts. Sintra, Portugal, 1-3 October 2019*. Brno, Lexical Computing, p. 10.

SINCLAIR, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London, Routledge, 212 p.

STEIN, J.; URDANG, L. 1966. *The Random House Dictionary of the English Language, The Unabridged Edition*. New York, NY, Random House, xxxii + 2059 p.

TONO, Y. 2000. On the effects of different types of electronic dictionary interfaces on L2 learners' reference behaviour in productive/receptive tasks. In: Heid, Ulrich, Stefan Evert, Egbert Lehmann & Christian Rohrer (eds). *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*. Stuttgart, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, p. 855-861.

TONO, Y. 2011. Application of eye-tracking in EFL learners' dictionary look-up process research. *International Journal of Lexicography*, **24**(1):124-153. https://doi.org/10.1093/ijl/ecq043

URDANG, L. 1966. The systems designs and devices used to process *The Random House Dictionary of the English Language*. *Computers and the Humanities*, **1**(2):31-33. https://doi.org/10.1007/BF00138355

URDANG, L. 1984. A lexicographer's adventures in computing. *Dictionaries: Journal of The Dictionary Society of North America*, **6**:150-165. https://doi.org/10.1353/dic.1984.0006

VAN GHEEL, J. 1652. *Vocabularium Latinum, Hispanicum, e Congense. Ad Usum Missionariorû transmittendorû ad Regni Congo Missiones*. Rome, Biblioteca Nazionale Centrale di Roma, fondi minori 1896, manuscript varia 274.

VAN WYK, E.B. 1995. Linguistic assumptions and lexicographical traditions in the African languages. *Lexikos*, **5**:82-96. https://doi.org/10.5788/5-1-1057

VARANTOLA, K. 2002. Use and usability of dictionaries: Common sense and context sensibility? In: Corréard, Marie-Hélène (ed.). *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. [s.l.], Euralex, p. 30-44.

WELKER, H.A. 2010. *Dictionary Use: A General Survey of Empirical Studies*. Brasília, [s.n.], 382 p.

WIEGAND, H.E. 1998. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin, Walter de Gruyter, xx + 1162 p.