Published in: Yearbook of the German cognitive linguistics association (2017), no. 5, pp. 107-118

Petra Storjohann

Cognitive descriptions in a corpus-based dictionary of German paronyms

https://doi.org/10.1515/gcla-2017–0008

Abstract: This paper discusses changes of lexicographic traditions with respect to approaches to meaning descriptions towards more cognitive perspectives. I will uncover how cognitive aspects can be incorporated into meaning descriptions based on corpus-driven analysis. The new German Online dictionary "Paronyme – Dynamisch im Kontrast" (Storjohann 2014; 2016) is concerned with easily confused words such as *effektiv*/*effizient*, *sensibel*/*sensitiv*. It is currently in the process of being developed and it aims at adopting a more conceptual and encyclopaedic approach to meaning by incorporating cognitive features. As a corpus-guided reference work it strives to adequately reflect ideas such as conceptual structure, categorisation and knowledge. Contrastive entries emphasise aspects of usage, comparing conceptual categories and indicate the (metonymic) mapping of knowledge. Adaptable access to lexicographic details and variable search options offer different foci and perspectives on linguistic information, and authentic examples reflect prototypical structures.

Some of the cognitive features are demonstrated with the help of examples. Firstly, I will outline how patterns of usage imply conceptual categories as central ideas instead of sufficiently logical criteria of semantic distinction. In this way, linguistic findings correlate better with how users conceptualise language. Secondly, it is pointed out how collocates are treated as family members and fillers in contexts. Thirdly, I will demonstrate how contextual structure and functions are included summarising referential information. Details are drawn from corpus data, they are usage-based linguistic patterns illustrating conversational interaction and semantic negotiations in contemporary public discourse. Finally, I will outline consultation routines which activate different facets of structural knowledge, e.g. through changes of the ordering of information or through the visualisation of semantic networks.

Keywords: cognitive lexicography, corpus semantics, paronyms, easily confused words, conceptual approach

Petra Storjohann: IDS Mannheim. E-mail: storjohann@ids-mannheim.de

1 Introduction

Lexicography has undergone dramatic changes over the past two decades. These mainly concern approaches to lexical analysis, the editorial process and the digitation/presentation of data. The relationship between semantic theory and practical lexicography has always been a difficult one (cf. Rundell 2012). When it comes to employing semantic foundations, it is above all the field of corpus linguistics that has left its mark on dictionary writing. Corpora and their tools have turned lexicography into a more objective and empirical trade which makes use of authentic language data. Lexicographers have also continually taken advantage of hypertextual opportunities to present lexical information in innovative ways, although their full potential has not been exploited, nor have users' needs been extensively considered (Müller-Spitzer 2014).

Cognitive linguistics, however, has not had any major impact on general dictionaries. In particular, the structuring of entries and the definition of senses are two areas where cognitive principles could be used to implement descriptions of conceptual structures and to show how meaning is construed or represented. As Ostermann (2015) points out, novel cognitive theories have been neither recognised nor successfully integrated into general English dictionaries. A few specialised frame-based English systems such as Pattern Dictionary of English Verbs (PDEV) or FrameNet[1] facilitate meaning descriptions with cognitive foundations (e.g. Fillmore 1976, 1977). As far as general monolingual dictionaries are concerned, both in English and in German, there is a complete lack of guiding cognitive elements (e.g. conceptualisation, construction, categorisation, representation) being incorporated into semantic descriptions with a theoretical foundation.

In this paper, it is argued that cognitive ideas can be successfully implemented in descriptions of meaning and the structuring of dictionary entries, and that these provide relevant information which primarily benefits users. In the following, the new German dictionary of commonly confused words "Paronyme-Dynamisch im Kontrast" (Storjohann 2016) is taken as an example that breaks with tradition by including central conceptual information and by representing both linguistic and encyclopaedic knowledge. Within the German context, it is a first attempt at a more cognitively infused lexicography calling for more realistic documentations of language and the way speakers perceive, conceptualise and

1 A related project in German is the German Frame-Semantic Online Lexicon GFOL (http://coerll.utexas.edu/frames/).

linguistically represent the world. For the purpose of illustration, some cognitive features will be demonstrated, particularly those emphasizing the interaction of details for more adequate depictions of flexible usage and contextual categorial implications.

2 Paronyms, dictionaries and language users' intuitions

Paronyms are easily confused words which regularly cause problems for both native speakers and language learners. As these lexical items often share morphological roots, they are similar with respect to sound, spelling and/or meaning, e.g. *sensibel*/*sensitiv*, *formell/formal/förmlich*, *Methode*/*Methodologie*/*Methodik*, *Elektrik*/*Elektronik*, *effektiv*/*effizient*.[2] Generally, such pairs/sets are not regarded as synonymous although corpus analyses suggest that some items undergo meaning change due to the rivalry between the words. Sometimes, they can develop synonymous notions and simply become lexical alternatives (cf. Storjohann 2015). In other cases, they remain similar in meaning but show subtle differences and restrictions in usage. Inevitably, situations of confusion arise when speakers' intuitions contradict information in existing reference works.

Today, speakers face a range of consultation options, from traditional print dictionaries to free online resources.[3] Unfortunately, these often do not provide satisfactory answers to questions about paronym behaviour, as the following quote from an online forum shows.

> bezüglich *sensitiv*/*sensibel*: „Ich hab zwar überall nach einer Definition dieser beiden Wörter gesucht, aber je mehr ich finde, desto irritierender ist es" (aus: http://depriforum.phpbb8.de/diskussionen-f16/sensitiv-sensibel-t1258.html).[4]

In online forums, people share their concerns about easily confused words. It is here, through the study of blogs, that detailed insights into the specific linguistic problems of users, their consultation behaviour and their needs can be gained.

2 For more examples see Schnörch (2015).

3 Most German e-dictionaries are copied or digitised versions of conventional reference books.

4 Translation: concerning sensitive/sensitive (or delicate, empathetic): I've looked for a definition of these two words everywhere, but the more I find, the more irritating it becomes.

Taking a closer look at resources such as Duden online, the lexicographic deficiencies of these become apparent. Users interested in the differences between *Elektrik*/*Elektronik* or *sensitiv*/*sensibel* find the following facts:

Elektrik: Gesamtheit einer elektrischen Ausstattung
Elektronik: Gesamtheit einer elektronischen Anlage oder Ausstattung

sensitiv: von übersteigerter Feinfühligkeit; überempfindlich
sensibel:
1) von besonderer Feinfühligkeit; empfindsam
2) empfindlich gegenüber Schmerzen und Reizen von außen; schmerzempfindlich
3) besonders viel Sorgfalt, Umsicht, Fingerspitzengefühl o. Ä. erfordernd, heikel

The entries of *Elektrik*/*Elektronik* are circular and "married with content from antiquated dictionaries – the type that define pedantic as 'of, pertaining to, or characteristic of a pedant'" (Rundell 2012: 74). The entries of *sensibel*/*sensitiv* mainly summarise synonyms. Users do not obtain sufficient details concerning their conceptual potential and contextual usage. They will miss information, for example, on semantic reference, relevant conceptual domains or categories, discourse structures and contextual situations. Who/what is specifically characterised as *sensibel*/*sensitiv* and in what kind of contextual circumstances? This question remains open. Consequently, online forums have turned into widely used social media sources where users consult the community for their linguistic problems. Frequently, they explain whole contextual situations in which their uncertainties occur. They seek information on lexical use, prototypical contexts, possible constructions, and conceptual as well as encyclopaedic issues. The answers of the language community are impressively diverse and revealing. As a matter of fact, speakers have good intuitions as to what linguistic and extra-linguistic information is required to form essential parts of authentic communication.

3 Paronyms from a cognitive perspective

Placing the user in focus, it is essential to strive for conceptual approaches and to document the interplay of lexical, structural and encyclopaedic knowledge in meaning descriptions. On the one hand, the implementation of a semantic structure and network that is closer to actual usage requires information on patterns of conceptualisation, on categories, reference and concrete lexical prototypes. On the other hand, we need to overcome a rigid, linear ordering of information and

strive for a more realistic representation of multi-dimensional facets of semantic configurations as we encounter them in language use. Only then, our descriptions can be closer to the current insights we have about the structure of the mental lexicon (cf. Ostermann 2015).

3.1 The dictionary "Paronyme – Dynamisch im Kontrast"

"Paronyme – Dynamisch im Kontrast" is an e-dictionary that breaks new ground by adopting a more conceptual and encyclopaedic approach to meaning by incorporating cognitive features.[5] It is currently in the process of being developed and includes conceptual, prototypical, and referential categorisation and a flexible structural access to knowledge. This dictionary does not follow sufficiently logical criteria of semantic distinction for its sense disambiguation. Instead, different patterns of usage and their underlying conceptual categories and prototypical realisations function as parameters of contextual distinction. These are then accessed flexibly via menu navigation.

As a quick guide, short paraphrases define characteristics of conceptual referential categories. Concerning the adjectival pair *sensibel*/*sensitiv*, relevant topic areas (or frame presentations) are given for each adjective. These are coded as "guide words"[6] together with a synonym (see Figure 1).

A large amount of knowledge about words, meanings and concepts is derived from experience and from the categories we construct, i.e. mentally represented frames or schemas. It is these categories (MATTERS/AFFAIRS, HUMAN, TECHNOLOGY, PROCEDURES, PLACES, MATERIAL, ORGANS, BEHAVIOUR, COGNITION, ANIMALS/PLANTS or COSMETICS) that justify a distinction of patterns and help to correlate situations of language use to different contexts. In the dictionary, these categories build up a quick contrastive guide and a concept-driven navigation structure (see Figure 1).

Through the visual explanations, it is possible to answer questions such as *Can German sensitiv be used together with places and materials*?, *Can an organ be sensitive* (German *sensibel)*? or *Can delicate plants be described as sensitiv or sensibel in German*? It is also possible to compare individual contexts of the two paronyms and quickly identify similarities and differences.

5 It will be published in the dictionary portal OWID (www.owid.de) in 2017.

6 Guide words are also used in *Cambridge International Dictionary of English*.

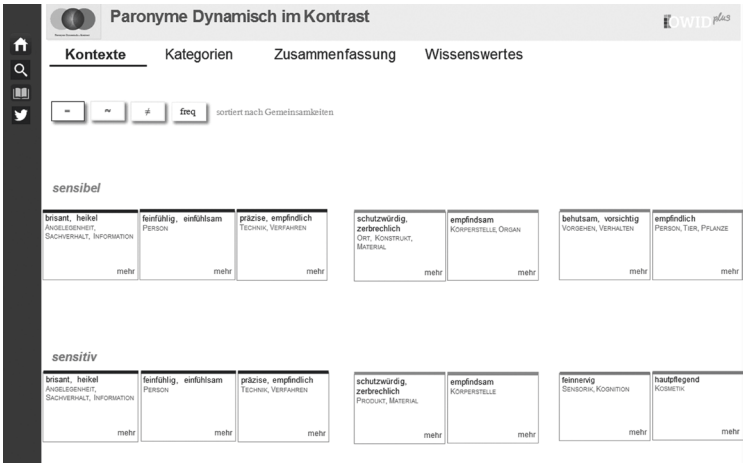

Figure 1: Default conceptual navigation structure (without menu options).

3.2 Fillers as lexical realisations

Users also have the option of consulting more detailed information on demand. Conceptual reference and encyclopaedic ideas are then explicitly integrated into the longer paraphrase. The relevant ontological category or domain is then specifically illustrated using lexical preferences, i.e. collocates. With a dynamic electronic display at hand, these are shown optionally, as a list of frequent and conventionalised contextual partners, introduced by *such as* underneath the definition (see figure 2).

In this approach, collocates are concrete lexical realisations (or fillers[7]) in specific contexts illustrating the referential category given in the definition.[8] For example, polysemous *sensibel* prototypically means something like 'delicate' or 'sensitive'. It is the conceptual background where the adjective refers to nouns functioning as non-human subjects or objects and denoting MATTERS OR AFFAIRS such as *data, subjects, information, questions* or *documents*. Similarly, German *sensitiv* also refers to nouns expressing the concepts of MATTERS OR AFFAIRS,

7 For verbs, which only make up only a small section of the dictionary, collocates serve as fillers in frame-like constructions. Collocates are then grouped into different sets (argument roles).

8 The linguistic analysis of corpus-driven collocates is also indicative evidence of distinct usage and senses. They are a primary source for lexicographers for deriving definitions and disambiguating meaning.

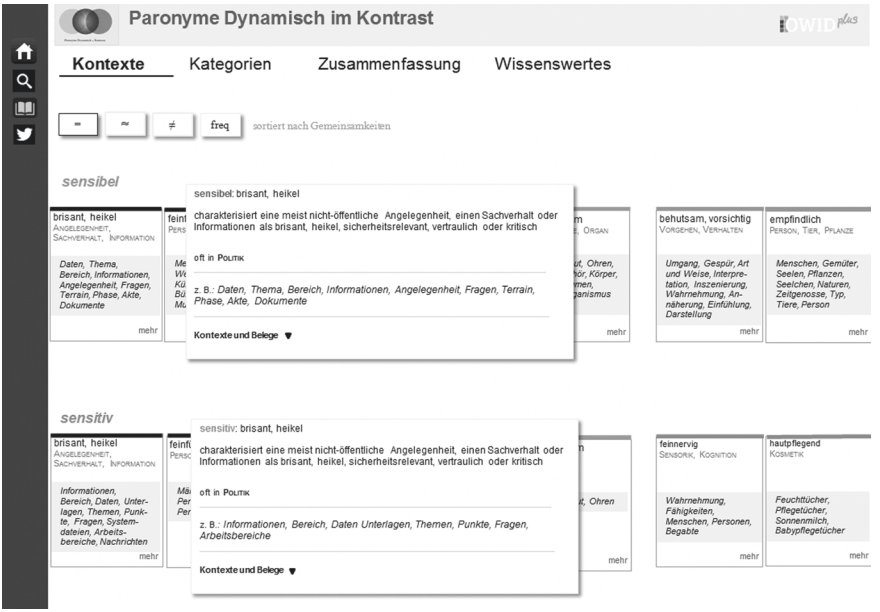

Figure 2: Long definition and prototypical realisation (fillers).

illustrated by *information*, *area*, *data*, *documents* or *questions*. Alternatively, both items can refer to the concept of TECHNOLOGICAL DEVICES or EQUIPMENT as 'highly precise and responsive': typically for *sensibel* these are *gas pedal*, *technology*, *seismograph*, *measuring instruments*, *instruments*. For *sensitiv* these are *robots*, *detectors, motion sensors* or *cameras.* In other contexts, they differ in terms of their conceptual referents, e.g. *sensibel* can be used to describe human behaviour or approaches (exemplified by *feeling*, *attempt*, *interpretation*) as careful and prudent. Plants, animals or people can also described as being *sensibel* in the sense of delicate or sensitive, being easily destroyed or offended. *Sensitiv* can be used to describe cosmetics (illustrated by *wet wipes*, *sun lotion*) as being gentle to delicate skin or the adjective denotes people as having a special sensory or cognitive ability.

The lexical representations are not intuition-based examples but statistically significant occurrences provided by corpus instances (see section 4). In essence, they are prototypical domain elements and structured mental representations of human experience. They shed light on strong affinities to constructions and contextual preferences, and they point to properties correlating with aspects of meaning structure. With prototypical details, we have the possibility of handling

polysemous contexts in a way that "more faithfully reflects what corpus data tells us" (Rundell 2012: 82). For polysemous items, metonymous and metaphorical contexts are listed. These show cognitive processes in which conceptual elements motivate the configuration of another semantically related conceptual entity (cf. Kövecses & Csabi 2014).

3.3 The organisation of knowledge

By focusing on the needs of the user, we have learned that these vary considerably (cf. Storjohann 2016). Given this, a system of various options has been developed which enables us to configure different perspectives on the organisation of knowledge. In essence, this dictionary is an XML-based hypermedia resource. Its system is customisable and can adaptively generate and prioritise information for specific user groups.[9] With multi-functional specifications at hand, dictionary data can be individually "reshuffled" by setting different parameters during the consultation process. Consequently, focal points on conceptual structures change and different linguistic aspects are emphasised.

Firstly, as a default, the different instances of usage of each lexical item are established in relation to the individual contexts of the corresponding paronym item, with identical contexts first, followed by similar and dissimilar contexts. Through this, an instant overview of overlapping uses and differences is provided (figure 1 or 2). Secondly, depending on personal interest, users can also choose parameters for listing the different contexts first. Thirdly, as neither ordering necessarily corresponds to the frequency of occurrence in actual usage, all contexts can be shown according to their distribution in the corpus, so that the predominance or centrality of certain contexts can be seen. Fourthly, it is the user's decision to choose the ordering of paronym items and determine which one appears at the top of the entry. Finally, the menu options also include a visualisation of collocation profiles with behavioural networks and interactive functions (see figure 3).

Taking the conceptual categories as a starting point, their corresponding collocational representations can be studied contrastively. The denoted concepts that are commonly shared are in the centre, followed by similar and dissimilar concepts arranged separately to the left and right below. Each category, together with its individual lexical realisations, is exemplified by corpus instances. It is a simplified diagram with abstract concepts directly representable in a contrastive

9 There are also extra regular search options.

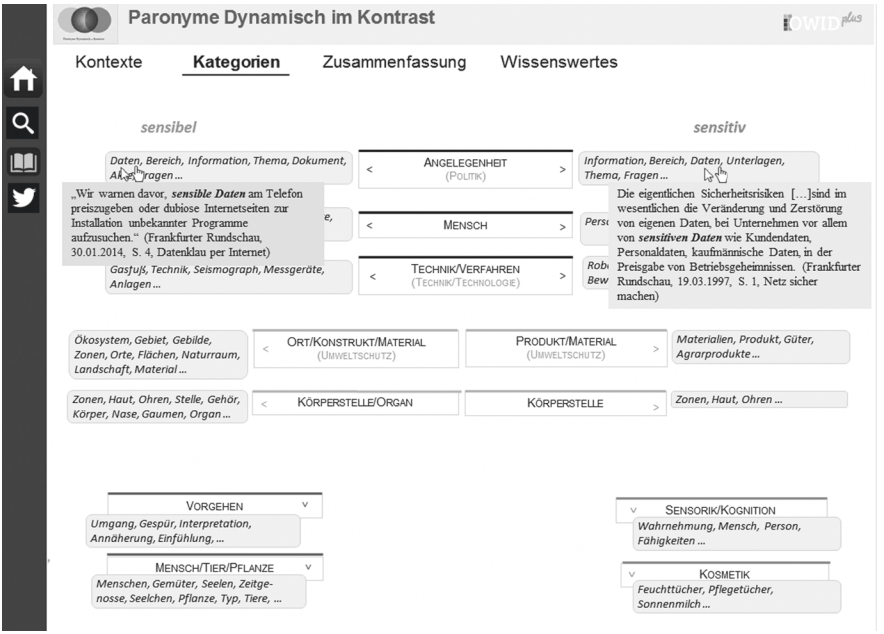

Figure 3: Visualisation of collocational profiles and interactive functions.

conceptual organisation. Overall, this e-dictionary exploits text- and hypertechnological possibilities and offers consultation routines by optionally generating different facets of structural knowledge.

4 Corpus Lexicography meets Cognitive Lexicography

The paronym dictionary bases its information on a comprehensive purpose-built corpus comprising 2.3 billion words.[10] As the subject of paronyms has not been revisited with empirical, data-driven methods, either in terms of semantic theory or in terms of practical lexicography, suitable corpus methods for contrastive investigation needed to be tested. Currently, complementary

10 For details see: http://www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html.

software-driven resources facilitating the search for similarity and difference are being exploited, each of which is based on the analysis and interpretation of contextual profiles, collocations and colligations, corresponding semantic roles and syntactic functions.[11] Corpus data reveals how meaning is constantly being negotiated in usage events and how communicative acts can create semantic rivalry or increase vagueness of easily confused words. Accordingly, variation and uncertainties arise from lexical similarity, sometimes leading to the adoption of new conceptual-semantic nuances. It is corpus-guided investigations that uncover discrepancies between conventionalised language use, speakers' intuition and traditional dictionary entries. They are essential in the tracing of regular, conventionalised or new semantic components. The analysis and interpretation of patterns shows that meaning is conceptualisation, constantly negotiated in usage. Aspects of discourse, domain, reference and ontological categorisation are mentally processed and stored as information on lexical use and meaning.

In the case of the paronym dictionary, linguistic and encyclopaedic details are drawn from corpus data and are included in usage-based linguistic patterns, illustrating conversational interaction and semantic negotiations in contemporary public discourse. Cognitive elements play an essential role when users confuse lexical items. This confusion is often not only related to formal similarities but also to conceptual closeness. Corpus-derived data allows for the search of minimal semantic differences and the integration of necessary encyclopaedic knowledge, information that is complementary to linguistic information and needed by users. While this is not news to cognitivists, lexicographers still have to learn how to integrate this insight into usable tools. Bridging the gap between corpus lexicography and cognitive paradigms is a slow but steady process (Gries 2006, Rundell 2012, Hanks 2013). Writing dictionaries should be informed by theoretical grounding and lexicographers should be linguistically aware corpus analysts. As Lew (2007: 221) points out, "let us hope that lexicographers will keep an open mind to developments in linguistics [...]".

5 Conclusion

So far, there is no corpus-assisted German reference guide describing commonly confused words empirically and enabling readers to find the correct contemporary

11 For verbs, these would be based on the extraction of complementation patterns.

usage.[12] The paronym dictionary is committed to overcoming the discrepancy between traditional practice and insights from language use. This necessarily means finding a way of educating users by showing how linguistic knowledge, encyclopaedic knowledge and human experience are inextricably linked. Given these goals, the dictionary breaks down the binary distinction of dictionary vs. encyclopaedia. Solutions to a number of lexicographical challenges were required. One aim was to bridge the gap between cognitive semantics and corpus lexicography by simultaneously considering users' needs. It was argued that cognitive aspects can successfully be incorporated into meaning descriptions based on corpus-driven analysis. Insights into collocational use and the interpretation of contexts can lead to the implementation of more abstract encyclopaedic or conceptual categories as central ideas. Together with concrete prototypical contextual realisation these replace circular definitions and uncommented lists of synonyms. Authentic examples reflect prototypical structures as manifested in discourse and in the mental lexicon.

In contrastive entries, the interaction between lexemes is emphasised. The dictionary strives to adequately reflect ideas such as conceptual structure, categorisation and knowledge. While Kövecses & Csábi (2014) argue that employing cognitive linguistics is a profitable theoretical underpinning for lexicographers, we favour the description in terms of cognitive principles as it predominantly embraces users' needs.

Only a digital resource is able to solve problems of strict macrostructural ordering. Indeed, "an online dictionary can be adapted to the needs of each dictionary user" (Kwary 2012: 35). Dynamic look-up options replace rigid structures. An adaptable access to lexicographical information has been suggested, where variable search options enable different foci and perspectives on linguistic information. In addition, the implementation of interactive collocation networks is a more onomasiological approach which offers an alternative access to language and knowledge structures relevant in actual usage events. The Paronymwörterbuch is a dynamic source of information where different users' interests will hopefully be met.

Dictionaries

Cambridge International Dictionary of English. 1995. Cambridge: CUP.
Duden Online. URL: www.dude-online.de. [accessed 10.05.2017]
FrameNet. URL: framenet.icsi.berkeley.edu/fndrupal/. [accessed 10.05.2017]

12 Intuition-based dictionaries include (Müller) 1973 and Pollmann & Wolk (2010).

German Frame-Semantic Online Lexicon GFOL. URL: http://coerll.utexas.edu/frames/. [accessed 10.05.2017]
Müller, Wolfgang. 1973. Leicht verwechselbare Wörter. Duden Taschenwörterbücher Vol. 17. Mannheim: Bibliographisches Institut.
OWID - Online Wortschatzinformationssystem des Deutschen. URL: www.owid.de. [accessed 10.05.2017]
Paronyme – Dynamisch im Kontrast. URL: www1.ids-mannheim.de/lexik/paronymwoerterbuch.html. [accessed 10.05.2017]
Pollmann, Christoph & Wolk, Ulrike. 2010. Wörterbuch der verwechselten Wörter. 1000 Zweifelsfälle verständlich erklärt. Stuttgart: Pons.
PDE – Pattern Dictionary of English Verbs. URL: http://www.pdev.org.uk [accessed 10.05.2017]

References

Fillmore, Charles J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280, 20–32.
Fillmore, Charles, J. 1977. The need for a frame semantics in linguistics. In Hans Karlgren (ed.), *Statistical Methods in Linguistics*, 5–29. Stockholm: Scriptor.
Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many senses of to run. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics Corpus-Based Approaches to Syntax and Lexis*, 57–100. Berlin, New York: de Gruyter.
Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge: MIT.
Kövecses, Zoltán & Csábi, Szilvia. 2014. Lexicography and cognitive linguistics. *Revista Española de Lingüística Aplicada* 27(1). 118–139.
Kwary, Deny Arnos. 2012. Adaptive hypermedia and user-oriented data for online dictionaries: A case study on an English dictionary of finance for Indonesian students. *International Journal of Lexicography* 25(1). 30–49.
Lew, Robert. 2007. Linguistic semantics and lexicography: A troubled relationship. In Małgorzata Fabiszak (ed.), *Language and Meaning. Cognitive and Functional Perpectives*, 217–224. Frankfurt: Peter Lang.
Müller-Spitzer, Carolin (ed.). 2014. *Using Online Dictionaries*. Berlin, New York: de Gruyter.
Ostermann, Carolin. 2015. *Cognitive lexicography. A New Approach to Lexicography Making Use of Cognitive Semantics*. Berlin & Boston: de Gruyter.
Rundell, Michael. 2012. It works in practice but will it work in theory? The unseasy relationship between lexicography and matters theoretical. In Ruth Vatvedt Fjeld, & Julie Matilde Torjusen (eds.), *Proceedings of the 15th EURALEX Congress,* 47–92 Oslo: University of Oslo. URL: www.euralex.org/elx_proceedings/.../pp47-92%20Rundell.pdf. (10 May, 2017.)
Schnörch, Ulrich. 2015. Wie viele Paronympaare gibt es eigentlich? Das Zusammenspiel aus korpuslinguistischen und redaktionellen Verfahren zur Ermittlung einer Paronymstichwortliste. *Sprachreport* 4. 16–26.
Storjohann, Petra. 2016. Vom Interesse am Gebrauch von Paronymen zur Notwendigkeit eines dynamischen Wörterbuchs. *Sprachreport* 4. 32–43.
Storjohann, Petra. 2015. Was ist der Unterschied zwischen sensitiv und sensibel? *Zeitschrift für Angewandte Linguistik* 62(1). 99–122.