

Empirical Approaches to German Paronyms

Petra Storjohann, Ulrich Schnörch
Institut für Deutsche Sprache Mannheim
storjohann@ids-mannheim.de, schnoerch@ids-mannheim.de

Abstract

German lexical items with similar or related morphological roots and similar meaning potential are easily confused by native speakers and language learners. These include so-called paronyms such as *effektiv/effizient*, *sensitive/sensibel*, *formell/formal/förmlich*. Although these are generally not regarded as synonyms, empirical studies suggest that in some cases items of a paronym set have undergone meaning change and developed synonymous notions. In other cases, they remain similar in meaning, but show subtle differences in definition and restrictions of usage. Whereas the treatment of synonyms has received attention from corpus-linguists (cf. Partington 1998; Taylor 2003), the subject of paronyms has not been revisited with empirical, data-driven methods neither in terms of semantic theory nor in terms of practical lexicography. As a consequence, we also need to search for suitable corpus methods for detailed semantic investigation. Lexicographically, some German paronyms have been documented in printed dictionaries (e.g. Müller 1973; Pollmann & Wolk 2010). However, there is no corpus-assisted reference guide describing paronyms empirically and enabling readers to find the correct contemporary usage. Therefore, solutions to some lexicographic challenges are required.

Keywords: paronyms; synonyms; easily confused words; collocation profile

1 Introduction

This paper presents a new lexicographic project studying easily confusable words in language use and employing a data-driven approach to the investigation of German paronyms. Although there is a large spectrum of definitions, paronyms are generally referred to as lexical items with both related or similar morphological roots as well as slight morphological difference such as suffixes. But they are not only linked to another by similarity of form and/or sound but also have a similar semantic potential and are hence commonly confused for one another. Alternatively, it is one of the items of a paronymic pair that is commonly misused both by native speakers and learners, respectively. Examples of such items include for example *effektiv/effizient* (*effective/efficient*), *sensitive/sensibel* (*sensible/sensitive*), *Method/Methodik/Methodologie* (*method/methodology*), *formell/formal/förmlich* (*formal*). They have generally not been regarded as synonyms (cf. Lăzărescu 1995, 1999). However, first empirical studies suggest that in a number of cases items of a paronym set have undergone meaning change and developed synonymous notions (e.g. Storjohann 2013). It is therefore argued here that a sharp distinction between paro-

nymy and synonymy is not always justified. In other cases, they remain similar in meaning, but can show restrictions of usage and subtle differences in definition.

Whereas the treatment of synonyms has received attention from corpus-linguists (cf. Partington 1998, Taylor 2003), the subject of paronymy has not been revisited with empirical, data-driven methods neither in terms of semantic theory nor in terms of practical lexicography. Lexicographically, some German paronyms have been documented in printed dictionaries (Müller 1973; Pollmann & Wolk 2010), although not systematically. However, there is no corpus-guided reference guide describing paronym sets empirically enabling readers to find the correct usage of such lexical items. So overall, paronyms should to be addressed from new perspectives. Firstly, the phenomenon has not been accounted for comprehensively in linguistic theory. Secondly, from a corpus linguistic view, we need to search for suitable corpus methods for detailed semantic investigation. Vachková & Belica (2009) propose the comparison of collocation profiles. They suggest a data-driven method to analyse lexical usage of near-synonyms with self-organising feature maps (SOMs). In this paper, we argue that this might prove a suitable method for the treatment of paronyms too, as it provides insight into both, semantic overlap and differences and it provides instant access to contrastive patterns by examining the concrete collocational behaviour of two items under investigation. Finally, solutions to some lexicographic challenges are required.

2 Linguistic Treatment of Paronyms

As Hausmann (1990) points out the subject of paronymy has mainly been approached linguistically from typological, language contrastive perspectives, particularly in the area of translation studies. These focus on this lexical relation exclusively from a language learner's view. Depending on different parameters, paronyms have been defined differently, covering items such as false friends, homophones, homographs or even cognates (cf. Bußmann 2002). In this paper, paronymy is broadly understood as a lexical relation between two or more items within one language, which are semantically related, have a similar or identical root and which are similar in form and sound but show a slight morphological difference (see morphemic paronyms/paronyms proper in Bolshakov & Gelbukh 2003: 199). Some of the most stringent terminology and a classification model have been proposed by Lăzărescu (1995, 1999) treating paronyms exclusively from a L2-learner's perspective.

Generally, Lăzărescu (1995) distinguishes between phonetically- and orthographically-based lexical confusion (e.g. *Föhn/Fön*), lexicologically-related terms (e.g. *Schiffahrt/Schiffsfahrt*, *Kindbett/Kinderbett*) and grammatically-based items such as *wohnen/bewohnen* or *dort/dortig*. Another class of paronyms exhibits specific semantic-stylistic features, e.g. as between *Vatermord/Vatermörder*, *Etikett/Etikette*). All these share their potential to be commonly confused in context and then to cause confusion in text reception or production due to similar phonemic representation as well as similarities of form and meaning. Most of these examples are far less likely to be mixed-up by native speakers. The importan-

ce of paronyms is based on the assumption that these items play a vital role for users in the process of second language acquisition and foreign language communication in order to avoid misunderstandings (Làzàrescu 1999). Confusing paronyms is sometimes regarded a violation of semantic correctness. Prescriptive analysts favour semantic correction and the avoidance of such mishaps and argue that paronyms “are important for poorly educated native speakers and for foreigners” (Bolshakov & Gelbukh 2003: 199). Indeed, the alleged misuse of morphologically and semantically similar words also cause linguistic uncertainties for native speakers, as numerous language-related Internet blogs show.

Language learners and native speakers too share their concern of correct language use with the language community as demonstrated by threads and their opening questions such as “Was ist der genaue Unterschied zwischen *effektiv* und *effizient*?” (What is the exact difference between *effektiv* and *effizient*), (see for instance <http://www.gutefrage.net/frage/was-ist-der-genaue-unterschied-zwischen-effektiv-und-effizient>). The answers of the blog community are impressively diverse.

So far, there is no semantic account encompassing different perspectives on the phenomenon and no satisfactory lexical focus on this relation that goes beyond Làzàrescu’s categorisation system and that comprises semantic, diachronic and cognitive aspects. Effectively, there is a large interest in easily confused words from both learners and native users alike, but we lack an empirical treatment and full theoretical account of paronymy in general. So far, the question of what constitutes a relation of paronymy has not been satisfactorily been answered as besides lexical features it also involves cognitive aspects. Furthermore, we have no suitable, user-friendly, appropriate dictionaries documenting paronym behavior (see section 4). Hence, there are no widely tested methods that have proved suitable for semantic analysis of such words. To be able to derive conclusion and to develop hypotheses, it is suggested to work with corpus-driven procedures to examine paronyms closer. With the help of corpora and innovative tools it is possible revisit paronymy and also to open up new research questions. Since it is possible to analyse language use synchronically as well as diachronically and observe gradual meaning change even over a short period of time it is possible to detect slight semantic shifts or nuances and also to determine the degree of semantic overlap between similar lexical items both quantitatively and qualitatively. It is argued (see also Storjohann 2013) that in some cases there is no semantic violation when paronyms are being confused. The meanings of typically confused words are more freely exposed to semantic negotiation. Following a descriptive empirical view, the semantics of some paronymic lexical items have adopted new semantic aspects and undergone meaning changes that are observable as regular patterns in a corpus and not as single misused occurrences. Overall, corpus-driven research on paronymy demands a more differentiated look at the phenomenon than has previously been offered. Empirically-based investigations of paronyms can also provide valuable insights into cognitive aspects and the exact circumstances under which two items are being confused as well as possible principles of language change of conceptually associated terms.

3 Corpus-Linguistic Approaches to Paronyms

Currently, researchers face a range of techniques that can be incorporated into the analysis of texts, such as eliciting co-occurrences, extracting keyword lists, investigating concordance and analysing dispersion and frequency of words or patterns. Methodologically, it is advantageous to use corpus tools that are able to provide good access to patterns and structures of lexical use by exploring co-occurrences. Exploring collocational patterns and other syntagmatic patterns has become an established procedure in order to describe the contextual behaviour of a word in empirical lexical semantics and in lexicography. It has also become an established tradition within corpus-linguistics to lead researchers more towards difference-driven analysis (cf. Taylor 2013). However, the examination of paronym sets necessarily incorporates contrastive meaning analyses including the study of similarities too. Therefore, suitable methods should be capable of measuring semantic similarity or distance by contrasting collocation profiles pairwise to systematically uncover overlap and differences in terms of contextual behaviour.

Storjohann (2013) conducted a contrastive analysis of paronymic items on two sets of data. A semantic study of *effizient/effektiv* has been carried out on the basis of a large newspaper corpus, mainly by investigating their collocations. The data used in the analysis consists of 2.7 billion words (cf. Storjohann 2005) and it has been the basis of the lexicographic project *ellexiko* (www.owid.de/wb/ellexiko/glossar/ellexiko-Korpus.html) at the Instiut für Deutsche Sprache. In addition, self-organising feature maps (SOMs) (cf. Kohonen 1990; Keibel & Belica 2007) have been employed which offer visual representation of topographic profiles of the involved lexical items and which complement the collocation analysis. This procedure is implemented into a corpus-linguistic research and development workbench called CCDB (Keibel & Belica 2007). The CCDB is a database containing numerous static lexical co-occurrence profiles. It has been used to extract topographic profiles to break down unstructured collocation patterns and hence complex semantic properties (see Figure 1 and 2). On the basis of collocation profiles, semantic structures are analysed, clustered, visualized in a two-dimensional lattice reflecting different degrees of similarity between various words. As emphasized before, the comparison of paronyms implies the analysis of difference as much as the analysis of overlap. These are systematically being identified between items with overlapping collocation profiles (cf. Vachková & Belica 2009). Self-organising features maps cluster all those items such that proximity of the grid reflects semantic similarity between their semantic profiles. The more similar the colours of two neighbouring groups, the more similar are their collocation profiles although a strict separation is not suggested, as SOMs imply a continuum of semantic shades. The more their colours differ the more semantic differences can be found with regard to their uses (see Figure 1 and 2).

© Cyril Belica: Modelling Semantic Proximity - Self-Organizing Map (version: 0.32, init tau: 0.04, dist: u, iter: 10000)

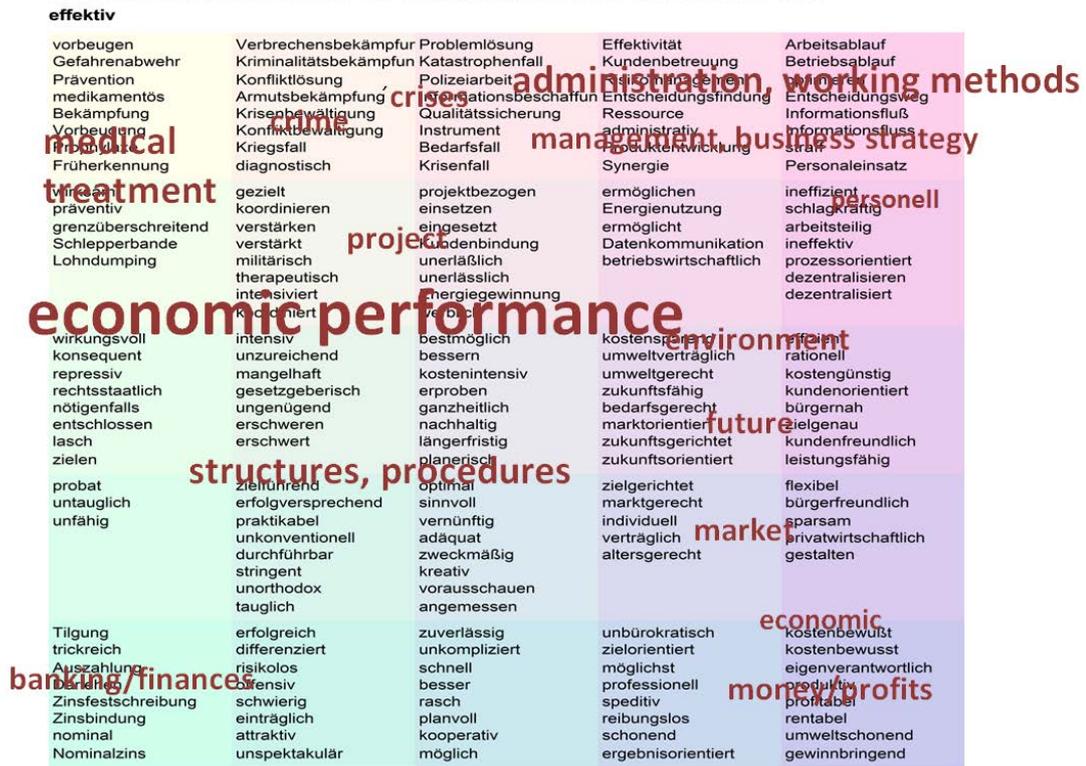


Figure 3: Topographic profiles of German *effektiv*.

© Cyril Belica: Modelling Semantic Proximity - Self-Organizing Map (version: 0.32, init tau: 0.04, dist: u, iter: 10000)

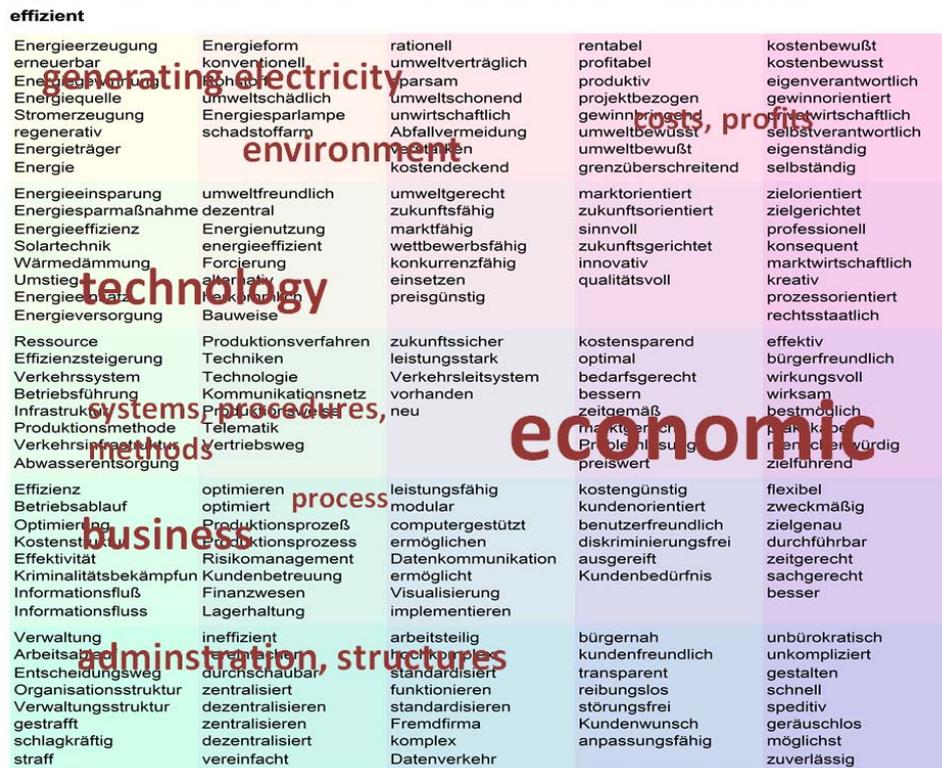


Figure 4: Topographic profiles of German *effizient*.

It is argued here, that the interpretation of such topographic feature maps could be the main entry point into data analysis of paronym behavior. It is a useful device for directing researchers to salient thematic domains associated with the individual terms. Through the process of abstraction a mental associations can be created by looking at individual squares and by moving from one square to the next (see Figure 1 and 2). As Vachková & Belica suggest:

Moving your focus forth and back, try to visualize the boundary where the initial sign eventually faded out, and where a notion of a new supersign entered your mind. Repeat for all corners and all directions. Try to assign each SOM square to at least one SOM supersign. (2009: 228-229)

Hence, a more abstract “supersigns” or superordinate concepts can be derived to categorise key semantic fields, clusters or domains. One major advantage of this procedure is the quick detection of thematic topics or domains in which the lexical items predominantly occur. For both lexical items, these thematic supersigns have been assigned and marked typographically. As a result, the interpretation of the findings as derived from supersigns or general concepts can then be directly compared, as summarised for example in Table 1.

effizient	effektiv
systems/procedures/structures administration costs/projects/economy	business/methods/work/management/personnel economic structures/performance
environment/generating electricity	fighting crime crisis management
technology	medical treatment/therapy
	environment/future
	banking & finance

Table 3: Semantic contexts/domains.

The domains in Table 1 are arranged according to their dominance in the underlying corpus. Semantic overlap between *effizient* and *effektiv* can be found in thematic domains where supersigns refer to the notion of business and administration and where both terms characterise methods, structures, procedures and issues of management. These are the discourse areas where ample evidence of synonymous contexts is provided in written German corpora (see examples 1 -3, for more examples see Storjohann 2013).

1. Zwar besteht auch in Hamburg die Vorschrift, dass ein Toter höchstens 36 Stunden in privaten Räumlichkeiten liegen darf und nach 14 Tagen begraben sein muss. „Aber Fakt ist, dass es eine weitere gesetzliche Reglementierung gar nicht gibt“, sagt Hillermann. Der sture Ablauf, der fast immer eingehalten wird, sei vielmehr das Ergebnis *effektiver* Arbeitsteilung von Spitälern, Bestattern und den Friedhofsbetreibern. Fast alle größeren Bestattungsunternehmen pflegen in

Deutschland eine *effiziente* Arbeitsteilung. (Die Zeit, 15.04.2004; Wie man in Deutschland begraben wird)

2. Im Jahr 2000 haben sich drei Fünftel der heimischen Gastbetriebe vom Dehoga abgespalten. [...] Der Hoga Rheinland, also der Hotel- und Gaststättenverband im nördlichen Rheinland-Pfalz, ist seinerzeit ausgetreten, weil er die Landes- und Bundesverbandsarbeit für nicht *effektiv* genug gehalten hat und die nicht üppig vorhandenen finanziellen Mittel lieber in regionale Arbeit und Projekte stecken wollte. Die Nordlichter warfen dem Dehoga zu wenig *effiziente* Arbeit vor. Wie sehen Sie das? (Rhein-Zeitung, 05.11.2004; Gastgeber wieder unter einem Dach)

Differences can be found for *effizient* being used as an attribute to characterise equipment, instruments and technological developments as well as being used to specify types and the use of energy resources. *Effektiv*, on the other hand, is more likely associated with the contextual domain of crime and crisis management. It is also attributed to characterise medical treatment and means of saving the climate. Furthermore, it is used within the context of banking and financing. Although, total semantic exclusion is not suggested, within these contexts both terms are far less likely being synonymously. As a result, the self-organising feature maps help us to structure the “unordered” semantics of a lexical item in use. It provides us with necessary details such as semantic dominance and contextual preferences in terms of referential domains and discursive foci.

In a further step, self-organisation maps can be used to contrast patterns of usage between two lexical items such as *effektiv* and *effizient* by comparing them with all those items with which they share parts of their collocation profiles.¹ This means that it is capable of measuring semantic similarity or distance by contrasting typical contextual behavior pairwise. The procedure is referred to as CNS-model (Contrasting Near-Synonyms, cf. Belica 2001 ff) and it is implemented into the workbench CCDB too. This procedure allows linguists to compare and contrast two words visually according to salient collocational contexts. *Effektiv* and *effizient* are being contrasted with each other and with those items with which they share parts of their co-occurrence profiles. The feature maps arrange specific usage aspects that the items in question share and those they do not have in common (see Figure 3).

1 Compare for example corpus software which facilitate the comparison of collocates visually, e.g. Sketch-Engine's Word Sketch Difference (<https://www.sketchengine.co.uk/>)

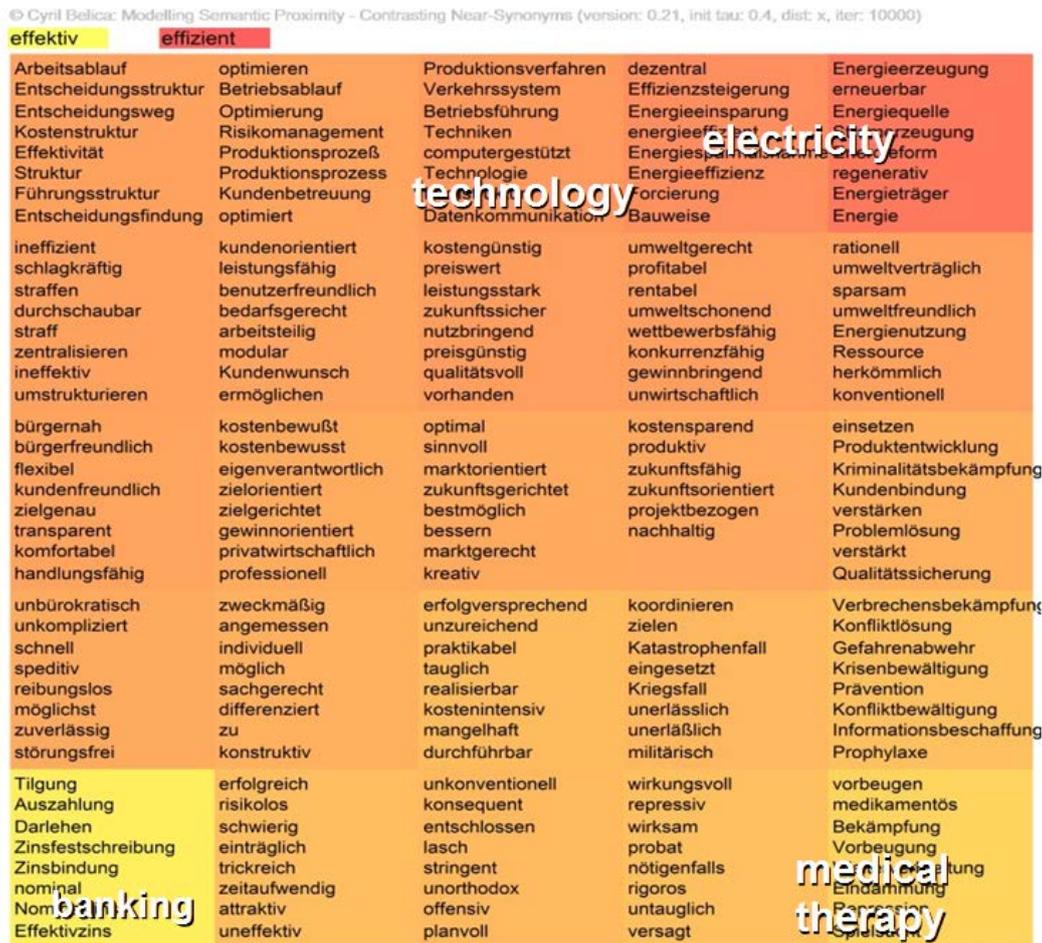


Figure 3: Contrasting German *effektiv* and *effizient* with SOM.

As indicated, the referential domains of ‘banking and financing’ (bottom left) and ‘medical treatment’ (bottom right) are preferably assigned to *effektiv*. Ways of generating electricity/energy and defining or specifying technology, however, are more likely a subject where the term *effizient* is preferably attested. In Figure 3, the subjective interpretation of the domains has been added to the map. Large parts of this feature map are shared by both terms. For the purpose being, these have not been further analysed also because a number of referential domains cannot always be clearly determined. Nevertheless, this feature map helps to validate the findings of conceptual similarities and differences as summarised in Table 1.

Generally, feature maps cannot serve as detailed lexicographic documentations to help users to be aware of “appropriate” or “false” usage. With respect to a paronym dictionary, more information as to concrete contextual usage is necessary. In the next stage, these domains or themes could, for example, be exemplified through statistically significant collocates extracted from an underlying corpus and assigned to corresponding discourses in order to illustrate specific preferences and restrictions (see Table 2). This process reveals for example, that one can modify procedures as processes (*Abläufe*, *Arbeitsabläufe*, *Arbeitsweise*, *Betriebsabläufe*), solutions (*Lösung*), structures (*Strukturen*), systems (*System*) and means/instruments (*Maßnahmen*) both as *effizient* and *effektiv* without implying much of a difference. It is within these contextual domains that similarity between the two terms is most evident

and examples of synonymous usage are being attested in the corpus (compare corpus examples 1 and 2). One does not typically characterise technology or the use of electricity/energy as *effektiv*. Here, the adjective *effizient* is being preferred. There are also efficient combustion engines (*Verbrennungsmotoren*) and efficient power stations (*Kraftwerke*). Other collocates on the topic of health and medical treatment include for example *therapy* (*Behandlungsmethoden, Therapie*) and *exercises for your back* (*Rückenübungen*) usually associated with *effective* but not with *efficient*. And terms referring to the banking and finance sector could be *interests, rates, return on capital* etc. These are modified in German by using *effektiv* (meaning real) exclusively, but never by *effizient*.

effizient		effektiv	
collocate	discourse domain	collocate	discourse domain
Abläufe, Betriebsabläufe, Arbeit, Arbeitsabläufe, Einsatz, Lösung, Strukturen, Verwaltungen, Maßnahmen, Arbeitsweise, Bewirtschaftung, Organisation, Wirtschaften, arbeiten, Aufgabenerfüllung, Bewirtschaftung, Methode, Abwicklung ...	systems procedures structures administration costs projects economy	arbeiten, gestalten, Maßnahme, Werbung, Methode, Maßnahme, Lösung, Strukturen, System, Arbeit, Verwaltung, Kommunikation, Kontrolle, Organisation, Arbeitsabläufe, Personaleinsatz, Zeitmanagement, organisieren, Controlling, Techniken, Strukturen ...	business methods work management personnel economic structures performance
Energie, Energieeinsatz, Energienutzung, Energieverwendung, Energieversorgung, Energieverbrauch, Heizen, Heizsysteme, Kohlekraftwerke, Stromnutzung, Stromerzeugung ...	environment generating electricity	bekämpfen, Mittel, Krisenmanagement, Strafverfolgung, schützen, Rechtsschutz, Selbstverteidigungstechniken, Verbrechensbekämpfung, Dopingbekämpfung, Polizeiarbeit, Überwachung ...	fighting crime crisis management
Motoren, Verbrennungsmotoren, Anlagen, Antriebstechnologie, Heizungen, Wärmepumpen, Kraftwerke, Geräte, Technologien, Wärmedämmung, ...	technology	Hochwasserschutz, Ressource, Klimaschutz, Wärmedämmung, Reduzierung	environment future
		Behandlungsmethoden, Ganzkörpertraining, Therapie, Rückenübungen, Prävention, Behandlung, Workout, Schmerztherapie, Beinmuskultraining, Behandlung ...	medical treatment therapy
		Jahreszins, Rendite, Zins, Zinssatz, Nominalzins	banking & finance

Table 4: Collocates in corresponding discursive domains.

As suggested by Vachková & Belica (2009), this approach to collocational patterning might be applicable to the lexicographic investigation of synonyms. It is argued that salient SOM features stimulate lexicographers' associative awareness and encourage guided mental imagery leading to valuable insights into both the word semantic structure and the process of discourse-based negotiation of lexical meaning (Vachková & Belica 2009: 239).

The notion of similarity has played a great role in lexicological areas, for instance in the corpus-linguistic investigation of sense relations by using collocational overlap to measure degrees of synonymy (cf. Partington 1998). Marková (2012), for example, puts forward examples of studies of German synonyms where she employed the CNS-model successfully. It is proposed here that consultations and interpretations of self-organisation feature maps might be a suitable approach to the analysis and semantic description of paronyms too where usage aspects that are shared and not shared can be uncovered. Feature maps can guide lexicographers to those contextual patterns where to look for further evidence for example through the analysis of collocations that can be attributed to specific thematic domains. Effectively, the chosen procedures result in a form of methodological triangulation comprising three different analytical stages: first, interpreting SOMs in order to associate domains, secondly, the CNS-Model to validate the previous interpretation and thirdly, collocation analysis to exemplify the given domains/topics contextually.

4 The Lexicographical Treatment

With regard to German, commonly confused words including some paronyms have been described in two printed reference books: Müller (1973) and Pollmann & Wolk (2010). Both are prescriptive documentations aiming at guiding users to the allegedly correct usage and describing a clear distinction between the items in question (see for example Figure 4).

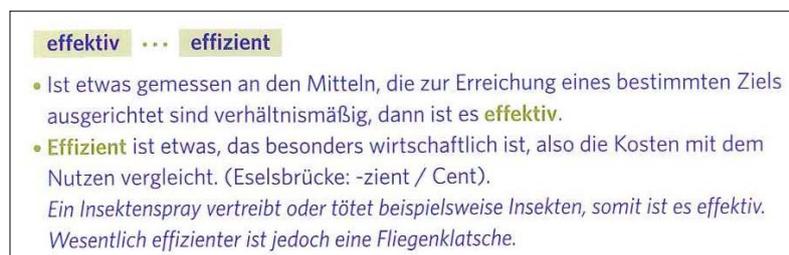


Figure 4: Dictionary entry *effektiv/effizient* in Pollmann & Wolk (2010).

The entries contain short meaning descriptions, occasional encyclopaedic comments and citations or examples. Moreover, some normative grammars and lexical studies concerning the didactics of normative language practice contain lists of some paronyms (e.g. Heringer 1989, 1995). Strictly normative language use is also propagated in wiktionary, a popular electronic resource which under an explicit

headline points out that confusion over the two words *effektiv* and *effizient* should be avoided (see Figure 5).

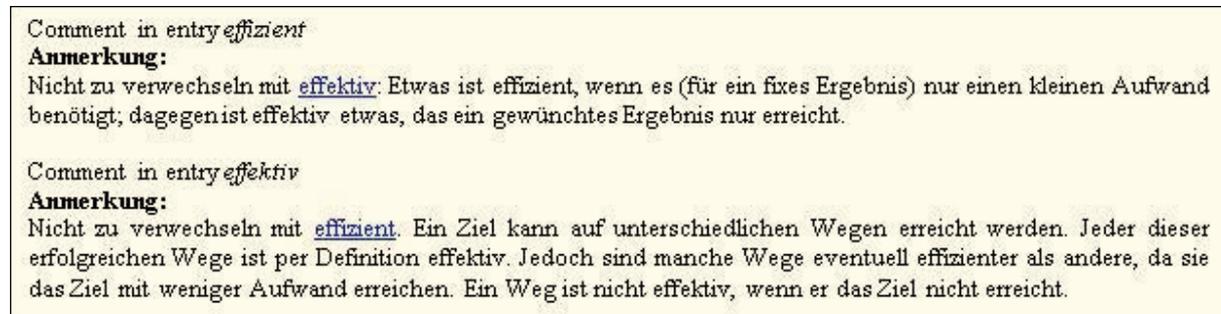


Figure 5: Comments regarding the misuse of *effektiv* and *effizient* in wiktionary.

From a lexicological point of view the remarks found there are questionable. However, they demonstrate that language users are aware of a potential conflict between *effektiv* and *effizient* and that misuse or confusion of this kind is a rather regular phenomenon that is to be avoided. Overall, in all reference guides findings are neither based on semantic examinations of current natural language in use nor on investigations of large data. Empirical corpus explorations open up the discrepancies to traditional descriptions. The usage restrictions that are documented in these reference books cannot be confirmed through corpus data. Entries lack collocational details referring to recurrent referential domains as for example illustrated in Figures 1 and 2. These provide essential information to users as to in which concrete contexts the corresponding adjectives might be more commonly found. As is the case for *effektiv/effizient*, strict usage lines cannot be sharply drawn which might have been expected intuitively. Conventional reference guides have so far focussed on the differences between commonly confused words. They entirely fail to explain existing similarities. In that respect the methodology is similar to most corpus-linguistic research. Corpus-assisted studies on semantically similar words have so far focussed on the differences between the individual items. As Pearce (2008: 21) points out there is a risk of “the privileging of differences over similarities. [...] the analyst is in danger of exaggerating the differences and overlooking similarities”. However, corpus studies also allow for the description of similarities which, on the one hand, might offer a deeper understanding why two words are regularly being confused and, on the other hand, it might indicate ongoing linguistic change worth documenting.

4.1 Challenges

A new project at the Institut für Deutsche Sprache will reopen the chapter on paronymy as a lexical as well as cognitive phenomenon. It will account for paronymy from a corpus-linguistic perspective and it will test methods that will hopefully prove suitable for semantic analysis of such words. It is hoped that data-driven investigations of paronyms can provide valuable insights into principles of

language change in semantically related lexical items. This could enable us to integrate paronymy into a wider theoretical framework. Part of this project is the compilation of the first corpus-assisted paronym dictionary which aims at guiding users descriptively through current usage and contexts. As an electronic resource it will provide adequate pairwise documentations combined with user-friendly navigation and search structures. In the near future, it will be an integral part of the German dictionary portal OWID (see: <http://www.owid.de/>).

From a lexicographic point of view, a number of challenges are encountered when documenting usage-based findings in a paronyms dictionary where users might demand definite answers for doubtful language situations. One central problem regards the interpretation and documentation of language change and normative restrictions. This is particularly relevant for pairs that are recorded as semantically distinct lexical items in traditional reference works and that have assimilated semantically over time due to common, allegedly “false” use. In some cases, corpus analyses signal tendencies that paronyms might have possibly turned into synonyms. Therefore, one of the major challenges of a corpus-based paronym dictionary is the interpretation of ambiguous data, especially paronym usage with a similar proportion between contexts with clear semantic difference between the terms and contexts exhibiting synonymous use. The lexicographic interpretation of such data requires a certain sensibility, as a specific conflict is expected to be encountered with corpus data. On the one hand, false language use caused by confusing paronyms needs prescriptive correction. On the other hand, gradual language change caused by frequent misuse of a certain lexical item needs descriptive documentation of contemporary language use. Above all, it should be able to explain semantic overlap and to sensitise users for continuous language change. Although this has not been studied, it is assumed that the expectations of users of such a dictionary have rather prescriptive notions due to their habits and their handling with other existing dictionaries. Therefore, the most challenging objective of this descriptive, usage-based dictionary certainly is to offer a reference guide that shows similarities and difference between paronymic items contrastively, including corpus samples, explanations and comments without neglecting to inform users about aspects of semantic overlap, gradual semantic changes, contextual vagueness, possible substitutability and at the same time still answering their look-up questions satisfactorily.

4.2 Presentation

As an e-dictionary the new German paronym guide can go far beyond the depth of information found in the two existing printed dictionaries. It will also have to consider different options with respect to navigation, visualisation, cross-referencing, linking and searching in order to exploit the possibilities of the electronic medium profitably and in order to create a user-friendly instrument. Traditional dictionary entries contain explanations of the formula “*effizient* is something that ...” or alternatively “if something is *effizient* than ...” (see Figure 4). The answer as to what exactly this something is can be found in contexts and collocates. Although at this point, no finite solution of the details of presentati-

on can be given, it is inarguably information on co-occurring patterns that most attention will be drawn to. Users can expect direct access to collocational patterns of two or more easily confused words together at the same time with their interpreted thematic domains in which both are likely to occur together with their preferences or their restrictions. Corpus samples will illustrate the information given. The depth of information could be realised as optional user-customised views. Overall, two aspects need further exploring. Firstly, the possibilities of the electronic medium still need to be examined and exploited to create a reliable and usable environment enabling users to make correct choices. Secondly, research on the users' needs and behaviour provides us with valuable insights (cf. Tarp 2008; Müller-Spitzer 2014) and these should be profitably incorporated into modes of presentation.

5 Conclusion

In this paper, a multifaceted approach to the study of regularly confused words in German has been discussed, suggesting a method of investigation that implies both differences and similarities between paronymic items. The phenomenon of paronymy has not been accounted for empirically. It has neither been reconsidered in recent linguistic theories and models nor in lexicographic practice. This gap will be closed by a new project which accounts for paronymy from a corpus-linguistic perspective where methods of investigation will be tested to find suitable tools for semantic analysis. With the help of the example *effective/effizient*, we have set out one possible way to implement different software-driven resources facilitating the search for similarity and difference. Corpus-assisted investigations of easily confused words and their usage over recent decades can provide valuable insight into principles of semantic shift. It is argued here, that such analyses might enable semanticists to integrate the phenomenon into a wider theoretical framework on the one hand and into appropriate lexicographic descriptions on the other hand.

6 References

- Belica, C. (1995). *Statistische Kollokationsanalyse und -clustering*. Korpuslinguistische Analysemethode. Mannheim: Institut für Deutsche Sprache.
- Belica, C. (2001ff). *Kookkurrenzdatenbank CCDB – V3.3. Eine korpuslinguistische Denk- und Experimentierplattform*. Mannheim: Institut für Deutsche Sprache. Accessed at <http://corpora.ids-mannheim.de/ccdb/> [27/3/2014].
- Belica, C. (2006). *Modellierung semantischer Nähe: Kontrastierung von nahen Synonymen*. Korpusanalytische Methode. Mannheim: Institut für Deutsche Sprache.
- Belica, C. (2011). *Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen*. In A. Abel, R. Zanin (eds.) *Korpora in Lehre und Forschung*. Freie Universität Bozen: Bozen University Press, pp. 155-178.

- Bolshakov, I.A., Gelbukh, A. (2003). Paronyms for accelerated correction of semantic errors. In *International Journal Information Theories & Applications*, 10, pp. 198-204.
- Bußman, H. (2002). *Lexikon der Sprachwissenschaft*. 3rd edition, Stuttgart: Kröner Verlag.
- ellexiko*. Accessed at <http://www.owid.de/wb/ellexiko/start.html> [27/3/2014].
- ellexiko-corpora*. Accessed at <http://www.owid.de/wb/ellexiko/glossar/ellexiko-Korpus.html> [27/3/2014].
- Hausmann, F.J. (1990). Das Wörterbuch der Homonyme, Homophone und Paronyme. In F.J. Hausmann, O. Reichmann, H.E. Wiegand (eds.) *Wörterbücher. Dictionaries. Dictionnaires. Vol. 2*. Berlin/New York: de Gruyter, pp. 1120-1125.
- Heringer, H.J. (1989). Lesen – lehren – lernen. Eine rezeptive Grammatik des Deutschen. Studienausgabe. Tübingen: Niemeyer.
- Heringer, H.J. (1995). Grammatik und Stil. Praktische Grammatik des Deutschen. Berlin: Cornelsen.
- Keibel, H., Cyril, B. (2007). CCDB. A Corpus-Linguistic Research and Development Workbench. In *Proceedings of the 4th Corpus Linguistics Conference, CL 2007, 27-30 July 2007*. University of Birmingham, UK.
- Kohonen, T. (1990). The Self-Organizing Map. New Concepts in Computer Science. In *Proceedings of the IEEE*, Vol. 78 (9), pp. 1464-1480.
- Lăzărescu, I. (1995). Deutsche Paronyme. In *Grazer Linguistische Studien* 43, pp. 85-93.
- Marková, V. (2012). Synonyme unter dem Mikroskop. Eine korpuslinguistische Studie. Tübingen: Gunter Narr.
- Müller, W. (1973). *Leicht verwechselbare Wörter. Duden Taschenwörterbücher*, Vol. 17. Mannheim: Bibliographisches Institut.
- Müller-Spitzer, C. (ed.) (2014). *Using Online Dictionaries*. Berlin/New York: de Gruyter. (Lexicographica: Series Maior 145).
- OWID Online-Wortschatz-Informationssystem Deutsch*. Accessed at: <http://www.owid.de/> [27/3/2014].
- Partington, A. (1998). *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.
- Pearce, M. (2008). Investigating the Collocational Behaviour of Man and Woman in the BNC using Sketch Engine. In *Corpora* 3(1), pp. 1-29.
- Pollmann, Ch., Wolk, U. (2010). *Wörterbuch der verwechselten Wörter*. Stuttgart: Pons.
- SketchEngine*. Accessed at: <https://www.sketchengine.co.uk/> [27/3/2014].
- Storjohann, P. (2005). Das ellexiko-Korpus. Aufbau und Zusammensetzung. In U. Haß (ed.) *Grundfragen der elektronischen Lexikographie. ellexiko – das Online-Informationssystem zum deutschen Wortschatz*. (Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter, pp. 55-70
- Storjohann, P. (2013). Korpuslinguistische und lexikografische Ansätze zur Beschreibung deutscher Paronyme. In H. Scheuringer, D. Sava (eds.) *Im Dienste des Wortes. Lexikologische und lexikografische Streifzüge*. Festschrift für Ioan Lăzărescu. (Reihe: Forschung zur deutschen Sprache in Mittel-, Ost- und Südosteuropa Vol. 3). Passau: Stutz, pp. 401-418.
- Vachková, M., Belica, C. (2009). Self-Organizing Lexical Feature Maps. Semiotic Interpretation and Possible Application in Lexicography. In *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*, 13(2), pp. 223-260.
- Wiktionary. Das freie Wörterbuch*. Accessed at: <http://de.wiktionary.org> [27/3/2014].
- Tarp, S. (2008). Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. Tübingen: Niemeyer.
- Taylor, Ch. (2013). Searching for Similarities Using Corpus-assisted Discourse Studies. In *Corpora*, 8, pp. 81-113.
- Taylor, J. (2003). Near synonyms as coextensive categories: 'Tall' and 'high' revisited. In *Languages Science*, 25, pp. 263-284.