Published in: Hardie, Andrew/Love, Robbie (eds.): Proceedings of the 7th International Corpus Linguistics Conference / Abstract Book in Lancaster CL2013. - Lancaster: UCREL, 2013. Pp. 275-277.

Lexical, corpus-methodological and lexicographic approaches to paronyms

Petra Storjohann
Institut für Deutsche Sprache Mannheim
storjohann@ids-mannheim.de

1 Introduction

German lexical items with related morphological roots and similar semantic potential are easily confused and misused both by native speakers and learners, respectively. Examples of so-called paronyms include *effektiv – effizient – effektvoll*, *scheinbar – anscheinend*, *formell – formal*. These are generally not regarded as synonyms. However, first 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 can 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 unfortunately 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 Lexicographically, some German paronyms have been documented in an outdated printed dictionary (Müller 1973). However, there is no corpus-guided reference guide describing paronym sets empirically enabling readers to find the correct usage of such lexical items. Overall, paronymy needs to be addressed from new perspectives. First, 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. 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 field of translation studies. However, we lack an empirical treatment and theoretical account of paronymy as a lexical phenomenon in general (cf. Làzàrescu 1995). Hence, there are no widely tested methods that 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 corpus procedures to examine paronyms closer. Empirically, corpus-driven investigations of paronyms can provide valuable insights into principles of language change in semantically related lexical items.

3 Corpus-linguistic approaches to paronyms

The examination of paronym sets necessarily incorporates contrastive meaning analyses. Methodologically, it is advantageous to use corpus tools that are not only able to analyse patterns by exploring co-occurrences. They should also be capable of measuring semantic similarity or distance by contrasting collocation profiles pairwise to systematically detect differences in terms of contextual behaviour.

One possibility could be the visual representation of topographic profiles of the involved lexical items and the comparison of those with self-organising feature maps (cf. Kohonen 1990; Keibel and Belica 2007) in order to contrast paronyms. Topographic profiles break down unstructured collocation patterns and hence complex semantic properties (see Figure 1).

Figure 4. Topographic profile of German *effektiv*

Furthermore, self-organisation maps can be used to contrast patterns of usage between two lexical items by comparing them with words which exhibit collocation profiles that are most similar to the two items in question (see Figure 2). This procedure referred to as CNS-model (Contrasting Near-Synonyms) has been developed and implemented in a German linguistic work bench (CCDB: co-occurrence database) by Belica (2001 ff).

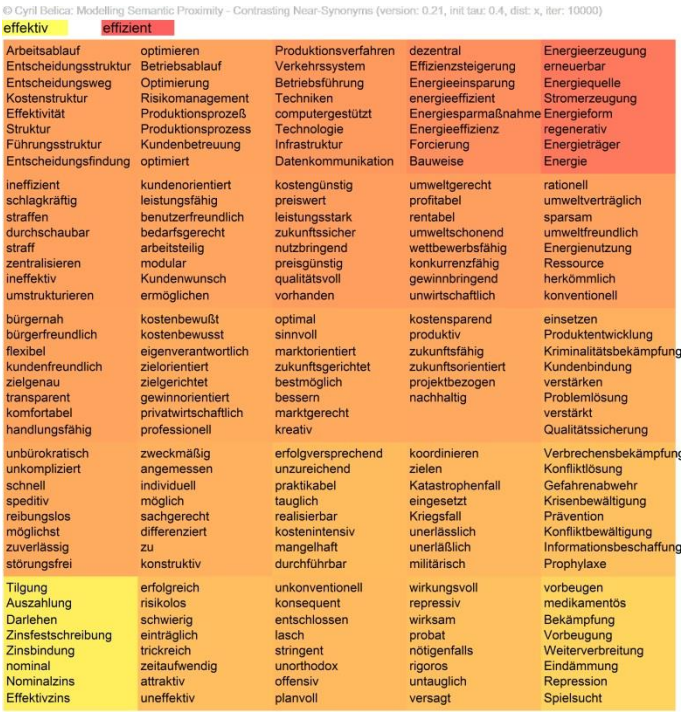

Figure 5. Contrasting German *effektiv –effizient* with SOM

The CCDB is used "for the study, development, and evaluation of methods for the data-driven exploration and modelling of language use" (Keibel and Belica 2007). SOMs arrange lexical items in two-dimensional lattices such that proximity on the grid reflects semantic similarity between collocation profiles. As suggested by Vachková and Belica (2009), this approach to collocational patterning might be applicable for lexicographic investigations of synonyms. The semantic properties of near-synonyms are contrasted with each other. Markova (2012), for example, puts forward examples of synonyms set which she investigated with the CNS-model successfully.

Consultations and interpretations of self-organisation feature maps might be a suitable approach to the analysis and semantic description of paronyms sets. It is argued that it might also be a practical corpus procedure for the examination of paronyms where usage aspects that are shared and not shared between easily confused words are detected.

4 Lexicographic challenges

From a lexicographic point of view, a number of challenges are encountered when documenting usage-based findings in a dictionary of paronyms which users generally expect to be rather prescriptive and where they 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. This assimilation process will have developed to different degrees among different paronym pairs/sets. In some cases, corpus analyses signal tendencies that paronyms might or might not possibly turn into synonyms.

Therefore, one of the major challenges of a corpus-based paronym dictionary is the lexicographic 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. One 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.

References

Belica, C. 1995. *Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analysemethode*. Institut für Deutsche Sprache: Mannheim.

Belica, C. 2001ff. *Kookkurrenzdatenbank CCDB – V3.3. Eine korpuslinguistische Denk- und Experimentierplattform*. Institut für Deutsche Sprache: Mannheim. Available online at http://corpora.ids-mannheim.de/ccdb/

Hausmann, F.J. 1990. "Das Wörterbuch der Homonyme, Homophone und Paronyme". In F.J. Hausmann, O. Reichmann and H.E. Wiegand (eds.) *Wörterbücher. Dictionaries. Dictionnaires*. vol. 2. Berlin/New York: de Gruyter, pp. 1120-1125.

Keibel, H. and Belica, C. 2007. "CCDB. A Corpus-Linguistic Research and Development Workbench". *Proceedings of the 4th Corpus Linguistics Conference* (CL 2007), Birmingham.

Kohonen, T. 1990. *The Self-Organizing Map. New Concepts in Computer Science*. Proc. Symp. in Honour of Jean-Claude Simon, Paris. AFCET, p.181-190.

Làzàrescu, I. 1995. "Deutsche Paronyme". *Grazer Linguistische Studien* 43, S. 85-93.

Müller, W. 1973. *Leicht verwechselbare Wörter. Duden Taschenwörterbücher Vol. 17*. Mannheim: Bibliographisches Institut.

Partington, A. S. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Vachkov , M. and Belica, C. 2009. “Self-Organizing Lexical Feature Maps. Semiotic Interpretation and Possible Application in Lexicography”. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 13/2: 223-260.

Taylor, J. 2003. “Near synonyms as coextensive categories:‘Tall’ and ‘high’ revisited”. *Languages Science*, 25: 263-284.