

## References

- Catford, J. C. (1983). "Insects are free": Reflections on Meaning in Linguistics. *Language Learning*, 3 (5), 13–32.
- Clear, J. (1994). I can't see the Sense in a Large Corpus. *Papers in Computational Lexicography. Complex '94* (pp. 33–48). Hungary, Budapest.
- Firth, J. R. (1957). *Papers in Linguistics 1934–1951*. Oxford: OUP.
- Hunston, S. & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the Lexical Grammar of English*. Amsterdam and Philadelphia: John Benjamin's.
- Kilgarriff, A., & Rundell, M. (2002). Lexical profiling software and its lexicographical applications – a case study. In A. Braasch & C. Povslen (Eds.), *Proceedings of the Tenth EURALEX International Congress* (pp. 807–818). Copenhagen, Denmark, August 13–17, 2002.
- Moon, R. (1998). *Fixed Expressions & Idioms in English*. Oxford: Clarendon Press.
- Palmer, H. E. (1933). *Second Interim Report on Collocations*. Tokyo: Kaitakusha.
- Sinclair, J. McH. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH. (Ed.). (1987). *Looking Up: An account of the COBUILD Project in Lexical Computing*. London: Collins.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamin's.
- Williams G. (2002). In search of representativity in specialised corpora: categorisation through collocation. *International Journal of Corpus Linguistics*, 7 (1), 43–64.

### Reviewer's address

Geoffrey C. Williams  
 Université de Bretagne Sud, France  
 4 rue Jean Zay  
 F 56100 LORIENT  
 France

\*\*\*\*\*

Bartsch, Sabine (2004). *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. Tübingen: Narr.

Reviewed by Nadja Nesselhauf (University of Heidelberg)

Although collocations have received much attention in linguistics recently, book-length studies of collocations in English are still fairly rare, and book-

length studies of this phenomenon adopting a corpus-based approach are rarer still. This book, the author's doctoral thesis, is an exception. On the basis of the British National Corpus, certain types of collocations are explored, and a methodology for dealing with collocations in a large computerized corpus is developed. The approach chosen for the study is referred to as "an integrated approach" (p. 111) by the author, as it combines qualitative and quantitative methods.

The book comprises eight chapters, of which two focus on the phenomenon of collocations as such, two on methodological issues, and two on the results of the corpus-based investigation (with Chapters 1 and 8 presenting a concise introduction and conclusion).

Chapter 2 provides an overview of the term 'collocation' as it has been used in linguistic research. It discusses its origins, its use by specialists in the area such as Firth and Hausmann, and some related concepts such as Coseriu's "lexical solidarities" and Porzig's "wesenhafte Beziehungen" ("intrinsic relations"). In addition, lexicographic treatments of collocations are surveyed as well as the status of collocations in different linguistic theories (Transformational Generative Grammar, dependency grammar, Head-Driven Phrase Structure Grammar, and Cognitive Grammar). In a final step, the different criteria employed in the various definitions of the term, such as the frequency of co-occurrence of items, constrained lexical selection and the degree of semantic cohesion, are isolated and considered individually.

In the third chapter, the potential defining criteria for collocations found in the literature are examined with respect to their applicability in a semi-automatic study of collocations in a large corpus. The definition developed on this basis regards collocations as "lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other" (p. 76). For practical purposes, the span size in which the lexical items have to co-occur in the corpus is determined in addition, and minimal requirements are established which co-occurring items have to meet according to three statistical tests.

The fourth chapter is devoted to the question of how potential collocations can be identified in corpora with statistical procedures. After some general theoretical remarks on statistical language modelling and the probabilistic nature of language, three different statistical tests commonly employed in the extraction of collocations, the t-test, MI-test and Chi-square-test, are described in detail and their potential and limitations discussed. A method that ensures the reliable identification of potential collocations is then developed, taking

into account the results of all three types of tests. It is found that for those co-occurring items for which both the MI-score and the t-score are significantly high, the probability that they constitute an actual collocation is close to 100%. The author stresses, however, that even such a sophisticated application of statistical methods has to be supplemented by manual analysis to filter out co-occurring items that are only potential but not actual collocations.

Chapter 5 gives a description of the British National Corpus, including its annotation, and of the procedures employed to select the items whose collocations were studied. The verbs studied in Chapter 6 were identified mainly with help of the *Collins English Thesaurus from A to Z* and the *WordNet* lexical data base. For the analysis presented in Chapter 7, the corpus was queried for part-of-speech patterns. A detailed description of the software used to extract potential collocations (Concordancer for Windows) and of the ways in which relevant data was extracted, stored, and accessed is also provided.

In Chapter 6, collocations involving verbal communication verbs (such as *admit*, *argue* and *tell*) are investigated. In particular, the adverbs collocating with such verbs are analysed, leading to the finding that most adverbs commonly collocate with only one or two, or at most three, verbs (*outright*, for example, collocates mostly with *reject*, sometimes with *deny* and very rarely with other verbs, cf. p. 150). If adverbs collocate with several verbs, these were found to be semantically related in most cases. An examination of the different metaphorical fields adverbs modifying verbal communication verbs come from is also presented. In a further approach, the semantic roles of arguments in structures involving verbal communication verbs are investigated with respect to their collocational behaviour. It is found that “collocational relations involving VCVs (verbal communication verbs) tend to hold predominantly between Message-Label and the Message-Content itself” (p. 174), such as in *ask a question*, *air views*, and *voice one’s opinion* on the one hand, and *admit defeat*, *claim responsibility*, and *declare truce* on the other. Collocational relations between such verbs and the items expressing roles of Speaker and Addressee are found to be rare. Very briefly, some delexical verb collocations (or support verb constructions) involving verbal communication nouns such as *make an announcement* are discussed in addition, the main result being a methodological one, namely “that it is necessary to combine different statistical algorithms in the corpus analysis of collocations to filter out different types of collocations.” (p. 171). A general finding, based on the investigations of both types of collocations discussed in this chapter is that “the number of dominant, i.e. very

frequent and statistically significant, collocational relations, is relatively small for most lexical items” (p. 178).

In Chapter 7, a different pattern is looked at, namely premodifier + adjectival past participle / desubstantival adjective + noun, as exemplified by *the guilt ridden middle class* and *a thick leafed plant*. A semantic classification of the adjectival elements in this pattern is provided. Among the groups that are established are ‘states of development / creation / preparation’ (e.g. in *captive bred animals* and *freshly baked rolls*) and ‘inalienably possessed body parts’ (e.g. in *a ruddy-faced person* and *a harshly boned face*). Then, the reason why the premodifier is obligatory in these constructions is investigated and found to lie in the overlap between the meanings of the adjectival element and the noun that it modifies (p. 191).

As this overview shows, the book has a very clear overall structure. Due to the fact that the author frequently repeats important points, the chapters can also for the most part be read individually. In particular Chapter 2 and Chapter 4 constitute fairly independent entities dealing with the phenomenon of collocations as such and with the statistical procedures of automatic collocation analysis. At the same time, both chapters are particularly noteworthy. The discussion of the concept of collocations is one of few that deal with the phenomenon comprehensively and one of the first overviews of the notion of collocations in different linguistic theories. The discussion of different statistical procedures for the extraction of collocations from a corpus is also exceptionally comprehensive, taking into consideration theoretical as well as practical issues.

As regards the actual study, the most innovative aspects are the investigation of the interrelation of semantic roles and collocations in Chapter 6 and the analysis of a fairly neglected pattern in Chapter 7. What would have been interesting in addition to the material presented in these two chapters is information on the overall results of the different investigations, i.e. how many types and tokens of each of the groups of collocations examined were found. The appendix is useful in that some of this information can be inferred from the numbers listed there. An exploration of the lexical combinatory possibilities of the pattern analysed in Chapter 7 would also have been of interest given that lexical constraint figures as a criterion in the definition of collocations; as things are, it is only asserted that lexical constraints do exist (p. 194).

Of the three aims set out at the beginning of the book and implicit in the title – to explore “how collocations may be identified in a large corpus of authentic language” and to investigate structural as well as functional properties of (some types of) collocations (p. 16) – it is the first that is achieved most

fully. Not only does the author develop a refined statistical approach to identify potential collocations in a corpus, she also develops a set of (other) criteria that can be employed to this end. A detailed outline of the technical aspects of collocation analysis after their extraction is also presented. Unfortunately, while the level of detail regarding the more technical aspects of the methodology is probably unrivalled (including information such as how exactly certain database programs were used), a few points in the procedure remain unclear, in particular with respect to the manual sifting of the results. One question that remains open, for example, is how the author determined which combinations are pragmatically and/or lexically constrained, i.e. how actual collocations were distinguished from the potential collocations yielded by the automatic extraction. Similarly, only a selection rather than the full set of verbal communication verbs chosen for the analysis in Chapter 6 are given (p. 142). What is more important, however, is the fact that the importance of manual analysis in addition to automatic analysis is recognized and an approach combining the two methods is adopted.

To sum up, the book can in many respects be considered an exploration and development of sophisticated semi-automatic methods for collocation extraction and is therefore of particular interest to researchers intending to perform phraseological analyses on the basis of a large corpus.

*Reviewer's address*

Nadja Nesselhauf  
Bahnhofstrasse 24  
76137 KARLSRUHE  
Germany

\*\*\*\*\*

Hans Lindquist and Christian Mair (Eds.) (2004). *Corpus Approaches to Grammaticalization in English*. (Studies in Corpus Linguistics 13). Amsterdam/Philadelphia: John Benjamins.

**Reviewed by Merja Kytö (Uppsala University)**

This is an appealing collection of papers, which brings together two topical themes, the corpus-based approach and the study of grammaticalization pro-