Idioms and Collocations is a collection of papers arising from a project in corpus linguistics at the Berlin Brandenburg Academy of Sciences. The papers are based around a list of 1000 German multi-word units (MWUs) selected from the DWDS (Digital Dictionary of the 20th/21st Century German Language) corpus. The project had three main goals: the establishment of techniques for the retrieval of a particular structure; the compilation of a database showing the linguistic properties of idioms and support verb constructions (SVCs) based on corpus evidence; and the provision of a research resource combining a dictionary and a corpus. The motivation for the project was that the authors felt that there was a lack of lexicographic consideration of German MWUs. While MWUs are difficult to classify neatly, the researchers did in fact find some patterns. As they indicate, however, it is hard for dictionaries to provide information on the flexibility of MWUs. The innovation in this book is the provision of corpus-based information on MWUs. The three areas on which the book is based are corpus linguistics, corpus-based lexicography and linguistics.

Research in this book focuses on SVCs and verb phrase idioms. SVCs are classified as ‘semantically fully compositional phrases’ (p. 8), such as *to take a risk*, and idioms are described as being semantically opaque, as in *to kick the bucket*. The introduction draws largely on Rosamund Moon’s treatment and classification of idioms.

The book is divided into two parts. The first, ‘Corpus, extraction and workbench’ (chapters 1 to 4), describes technical aspects of corpus searching and database creation in relation to idioms. The second, ‘Linguistic analysis’ (chapters 5 to 10), presents corpus-based studies on different linguistic features of MWUs.

Chapter 1, by Alexander Geyken, is entitled ‘The DWDS corpus: a reference corpus for the German language of the twentieth century’. It explains that one motivation for the DWDS project was that there was no adequate corpus or dictionary representative of twentieth century German language. Material for the corpus was taken from prose, verse and drama; newspapers; scientific publications; non-fiction (such as cookery books and advertisements); and transcriptions of speech recordings. East and West German texts are equally represented, but Austrian and Swiss texts are ‘deliberately under-represented’ (p. 30). Over two thirds of the corpus is publicly available from the DWDS website.
In chapter 2, ‘Classifying NVGs/FVGs in an interactive parsing process’, Alexander Geyken and Alexey Sokirko describe a semi-automatic parsing process for classifying verb-nominalisation constructions and the associated subclass of function verb constructions. The lexicographer can set up a filtering system to pre-classify such constructions, thus avoiding the need for random sampling. The results in this case show a high identification precision rate of between 97.3% and 99.7%.

Chapter 3, ‘Corpus queries’ by Axel Herold, addresses the creation of result sets for idiomatic expressions in a corpus. The goal is to describe grammatical authentic corpus examples of German idioms. The problem is that linguistic description is only possible after the result sets have been identified, and yet such a description is necessary in order to restrict the corpus search. Intuition is a fallible guide in linguistic description, with different native speakers, for example, often choosing different prepositions in an expression. The same problem of variants can occur in dictionaries, with one form given preference over another. Herold explains that there are a variety of techniques for measuring co-occurrence, but as yet it is still not easy to find idioms in a corpus. However, the researchers have developed a list of principles to be used when searching for idioms, among them keeping the search as simple as possible and being aware of all possible modifications (p. 61).

Chapter 4 presents ‘A lexicographic workbench for German collocations’ by Gerald Neumann, Fabian Körner and Christiane Fellbaum. This chapter describes the process of sorting and structuring corpora. The template feature allows users to search for information on particular idioms, including information on an idiom’s structure, variations, paradigmatic relationships and semantic field.

Chapter 5, ‘A corpus-based analysis of adjectival modification in German idioms’ by Katerina Stathi, takes us into the linguistic analysis part of the book. The chapter analyses attributive adjectives and participles. Examples are given of nouns literally modified by adjectives which relate to the idiom’s source domain (e.g., take the strong wind out of someone’s sails). This internal modification provides evidence that the underlying metaphor on which the idiom is based is still understandable to its users, as the adjective and noun match semantically. In external modification, the adjective modifies the entire idiom. Externally modifying adjectives do not have to collocate with the noun and often function as adverbs. (e.g. Die politische Entwicklung außerhalb Deutschlands – in China, Vietnam, Kambodscha, Persien und der Sowjetunion – hat dem Terrorismus
den ideologischen Wind aus den Segeln genommen. Lit. ‘The political development outside Germany ... has taken the ideological wind out of the sails’ = has weakened terrorism ideologically, p. 94). A third type of modification is what Stathi calls ‘intermediate modification’, applying to more abstract idioms. The fourth, still referring to adjectives, is called ‘conjunction modification’, while the fifth is metalinguistic modification, in which an adjective underscores the fact than an idiom is being used (an English example would be *bite the proverbial dust*). The findings in this chapter suggest that the same idiom may be used in either an analysable or a non-analysable way, focussing on the literal image or the abstract conception.

Chapter 6 is ‘Types of changes in idioms – some surprising results of corpus research’ by Elke Gehweiler, Iris Höser and Undine Kramer. This chapter highlights variations in form (e.g. prepositions) and meaning and makes the interesting point that ‘commercial idiom dictionaries are hardly ever based on corpus data’ (p. 109). According to these authors, some idioms may be polysemous, with related meanings derived from one original expression, while others (though rarer) may be homonymic, with unrelated meanings. A dictionary should reflect this, as well as degrees of transparency and opacity. Idioms can also change in meaning over time, or become so familiar that the original noun is omitted. (An English example of this would be *pull the other one.*)

The chapter reveals morphosyntactic changes in idioms made by users unaware of the idiom’s original etymology. The corpus is also used to identify new idioms. For example, *keine Geige spielen* (to play no fiddle = to be unimportant) is first attested in 1989 (p. 133). This work is necessary as new idioms emerge, and little research has yet been done in this field.

Chapter 7, ‘Meaning and use: a corpus-based case study of idiomatic MWUs’ by Christiane Hümmer, notes that idioms can be both arbitrary and motivated in their contextual behaviour. Idioms are portrayed as having three levels of meaning: literal, figurative (applied to individual constituents) and idiomatic (in which the MWU is taken as a whole). These may sometimes be related and may affect the way they behave in context.

Chapter 8 is “‘You fool her” doesn’t mean (that) “you conduct her behind the light”: (Dis)agglutination of the determiner in German idioms’, by Anna Firenze. It considers the use of articles in idioms, and in particular addresses agglutinated articles (e.g., vom = von dem) and the extent to which they are central to an idiom’s figurative meaning. The evidence in this chapter suggests that some idioms with contracted prepositions may not be frozen, and when articles are deleted, the meaning of the idiom becomes transparent.
In chapter 9, ‘Corpus-based investigations on German support verb constructions’ by Angelika Storrer, the online 2005 DWDS corpus is used to provide empirical evidence regarding the properties of SVCs. The study aims to distinguish stylistic and functional differences between SVCs, such as to take a risk, and base verbs, such as to risk. SVCs are shown to be semantically more specific than base verb constructions and to offer different ways of conveying information.

Chapter 10 is entitled ‘Argument selection and alternations in VP idioms’, and is written by Christiane Fellbaum. The study indicates how idioms differ syntactically as well as semantically from non-idiomatic language. This enables them to carry certain meanings which might not be expressed as succinctly in ordinary language.

This book provides a wealth of information regarding corpus searching and idiom properties. It is not for the novice, as some complicated concepts are explored. Even the acronym DWDS, although referred to on p. 2, was not explained until p. 23. In a minor hiccup, p. 43 refers to colours while the text is black and white. I had a query about the origin of one idiom in chapter 6: go to the dogs (p. 116) is referred to as having a Biblical origin. Having searched the OED and two English Bible versions (the New International Version and the King James Version), I cannot find this idiom as having a Biblical origin. Perhaps it is used in German Bible translations? I also found the term ‘idiomatic meaning’ in chapter 7 a little confusing.

Idioms and Collocations is probably most useful for those studying corpus linguistics, though those undertaking studies in the area of idiom properties would also benefit from the second part. Overall, I found it a challenging but thought-provoking read.

Review by Julia Miller
University of Adelaide