THOMAS HERBST AND KATRIN GÖTZ-VOTTELER

INTRODUCTION: THE MYSTERY OF COLLOCATION

The term *collocation* presents an almost prototypical example of the phenomenon of polysemy. In what one could call a rather neutral sense it is used to refer to any combination of two words that co-occur in language text. Other uses of the term identify more specific types of combination but differ as to what they consider to be the relevant criterion that makes up specificity. One type, the *sandy beaches*-type, refers to specificity in statistical terms of co-occurrence in the language, or in a corpus, – where statistical significance is not necessarily determined in terms of absolute frequency of co-occurrence but calculated on the basis of some sort of measure of mutual expectancy. In the second type, the *guilty conscience*-type, the combination is significant because it is established or institutionalized, to use a term common in word formation, and somehow unpredictable on the grounds of the meanings of the words. It is relatively obvious why the first type, referred to as “quantitative” by Hausmann and Blumenthal (2006, 3) should feature prominently in corpus linguistics, whereas the second, which Hausmann and Blumenthal (2006, 3) characterize as “essentiellement qualitative” (essentially qualitative) and in terms of “cooccurrence lexicale restreinte” (restricted lexical co-occurrence), has been a key concept of foreign language linguistics for a very long time.

That a foreign speaker of English who knows the noun *tea* will need to know (or be able to find out from a dictionary) that it can be qualified as *weak tea* but not as *feeble tea or light tea* or that the verb that accompanies *tea* to express the idea of preparation is *make* and not *cook or boil* is an insight that is reflected in Makkai’s (1972) classification of such combinations as “encoding idioms” and in Hausmann’s fundamental distinction between *Basis* (*tea*) and *Kollokator* (*collocate*), which is of particular relevance to foreign language teaching and foreign language lexicography.

While the original vagueness of the term as used by Firth can be seen as leading to the emergence of its different uses, there is also considerable overlap

---

Thomas Herbst and Katrin Götz-Votteler

between them. What is of crucial importance, however, is that both foreign language linguistics and corpus linguistics provide ample evidence for the fact that language text, or discourse, is not the result of some sort of application of grammatical rules followed by lexical insertion, or, in other words, that grammaticalness is not "the only restraint" in lexical choice (Sinclair 1991, 109). Rather, syntagmatic relations between lexical items in the sense of co-selection seem to play a crucial role in language. A programmatic title such as "Wortschatzlernen ist Kollokationslernen" (Hausmann 1984) is perfectly compatible with Barlow’s (2004, 213) statement that "collocations are the building blocks of language and are, in some sense, fundamental units of language in use", although they represent different uses of the term collocation. In other words, it might make sense to regard the sandy beaches-type of collocation and the guilty-conscience-type as two prototypical centres of the same phenomenon. Certainly, both can be subsumed under Sinclair’s (1991, 110) idiom principle, namely “that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.”

In fact, coming from a foreign language background one might ask to what extent a combination such as sandy beaches should be seen as a free combination of two lexical items at all since the fact that the concept of a sandy beach is expressed in German by a compound, Sandstrand, would generally be taken as a reason for considering this a single lexical choice. Since sandige Strände does not seemed to be used in German (although feinsandige Strände do occur), even what appears to be a free combination, such as sandy beaches, in English is in some way subject to usage or the level of Norm in Coseriu’s (1973) sense. It is interesting to see that the parallel between free combinations and compounds is already apparent in the writing of John Rupert Firth (1956/1968, 107), when he talks about “collocational compounds and collocational phrases in which common words appear” and gives examples such as safety match and safety first.

While phraseology in the past has concentrated on studying the relations between different lexical forms which often co-occur, the correlation between the expression of particular meanings in a particular way also deserves investigation. For instance, the time span comprising some 180 days can be referred to as half a year/halbes Jahr or six months/sechs Monate in English and German; nevertheless these different expressions are not really equivalent. In English, six months is definitely the preferred way of expressing that meaning (with 4009 instances for six months in the BNC as against 49 of half a year), whereas in German no such clear picture emerges. Units such as six months, which play a major role in bilingual lexicography, for instance, have been termed probabemes (Herbst/Klotz 2003 and forthcoming) and it can be argued that knowledge of probabemes is part of the competence of the speakers of a language. It is obvious

---

that the study of probabemes in different languages also cuts across the traditional distinction between words or lexemes on the one hand and combinations of words or collocations on the other, as the example of Sandstrand and sandy beach shows quite clearly. The fact that Sandstrand is lexicalised in German (and found in dictionaries), whereas sandy beach is commonly considered a free combination (which is not lemmatised in dictionaries) suggests that, in Sinclair’s terms, when talking about sandy beaches, speakers of German carry out a single lexical choice whereas speakers of English carry out two lexical choices. This obviously raises the fundamental question of the relationship between conceptualisation and lexical elements.

As far as such psycholinguistic or cognitive issues are concerned, it is very encouraging that within cognitive linguistics construction grammar “grew out of a concern to find a place for idiomatic expressions in the speaker’s knowledge of a grammar of their language” (Croft/Cruse 2004, 225). So far, construction grammar has not really offered any new insights into the nature of the different types and degrees of fixedness and variability exemplified in the volumes of the Oxford Dictionary of Current Idiomatic English, for example. However, in the long run, it might provide a framework which integrates insights into the idiomatic character of language in terms of the idiom principle into a comprehensive model of language which is compatible with valuable insights about the collocational and idiomatic nature of language provided by foreign language linguistics and corpus linguistics.

It was this range of issues that was in the focus of attention of a series of papers given at the Interdisziplinäres Zentrum für Lexikografie, Valenz- und Kollokationsforschung at the Friedrich-Alexander-Universität Erlangen-Nürnberg in the summer term of 2006, which provided the impetus for this volume. Franz Josef Hausmann’s (Erlangen) opening article offers an outline of his theory of collocation and the history of the term, referring mostly to its use in the description of French and English. Dirk Siepmann (Osnabrück) discusses various concepts of collocation with respect to the teaching of English and French to German students and gives an outline of the Bilexicon project. The aspect of collocation in foreign language teaching is covered in two further papers: Diana Lea (Oxford University Press) provides insights into the linguistic and lexicographical principles underlying the compilation of the Oxford Collocations Dictionary for Students of English, and Gaëtanelle Gilquin (Louvain) deals with the analysis of learner corpora in English and French. Göran Kjellmer’s (Göteborg/Gothenburg) contribution is an example of collocational analysis in corpus linguistics, focusing on constructions of the type a couple hours offering a diachronic perspective. Hans-Jörg Schmid’s (München) article focuses on an analysis of shell nouns with respect to Sinclair’s idiom principle in a construction grammar framework.

---

4 Compare Lea’s discussion of grey area (this volume).
5 See also Fillmore/Kay/O’Connor (1988) and Goldberg (2006)
6 Compare, for example, Granger (1998).
Works Cited

Dictionaries


Other Sources


−(this volume). “Die Kollokationen im Rahmen der Phraseologie – Systematische und historische Darstellung.”

Lea, Diana (this volume). “Making a Collocations Dictionary.”