Towards Collocational Webs for Presenting Collocations in Learners’ Dictionaries

Susanne Handl

Introduction

Researching collocations in a foreign language requires both a detailed consideration of native speakers’ habitual word combinations and an account of how those combinations can be made accessible to language learners. Whereas acquiring collocations in an L1 is a natural process based on constant exposure to language in context and co-text (see Handl and Graf, 2009) in the completely different situation of L2 acquisition, the teaching/learning environment and materials have to compensate for the lack of linguistic input. A major source of information for learners besides textbooks and teachers is the dictionary. Although a range of different dictionaries is available to help students cope with any linguistic task, it seems that on average they prefer a single book for everything. Thus, it would certainly be helpful, if as many collocations as possible could be integrated in such a dictionary.

In this chapter I propose a method that aims at improving the representation of collocations in advanced learners’ dictionaries in a number of ways. Drawing on the findings of a large-scale analysis of collocations (Handl, in preparation) in the British National Corpus (BNC; Oxford University, 2005), the present study explores a new multi-dimensional classification for working out in detail the criteria by which collocations can be selected for advanced learners’ dictionaries. Based on this, I present a method of display for collocation entries and also report on a pilot study evaluating this alternative.

Research basis

Theoretical background

Collocation has seen a wide range of classifications, from simple binary approaches (Firth, 1957c) and gradual classifications (Benson, Benson...
and Ilson, 1986a; Carter, 1998) to a prototypical account of the category of collocation (Schmid, 2003: 249). As collocation is a pervasive phenomenon that escapes clear-cut categorization, the prototype model and recent approaches instigated by corpus linguistics (Sinclair, 1996; Stubbs, 2002) seem most promising. They integrate statistical aspects of frequency to delimit – at least approximately – the scope of collocations. The fuzzy boundaries, however, are an obstacle to lexicographic description of collocations. That is the reason why, in the so-called significance-oriented approach, researchers like Hausmann (1984), Benson (1985) and Klotz (2000) promote clear-cut boundaries. They classify collocations typologically mainly on the basis of word classes and the semantic relation between base and collocator. To my mind, however, this approach misses out on incorporating psycholinguistic and cognitive factors that can be captured with the latest trends in corpus linguistics.

For a more detailed working definition, the basic notion of collocation as ‘the occurrence of two or more words within a short space of each other in a text’ (Sinclair, 1991: 170) needs to be enlarged with a set of further criteria. First, the frequency of such combinations and the respective co-occurrence measures can be used to rank significant collocations. Another factor that points towards their classification on a continuum is the fact that collocations are more or less restricted in their choice of partners, and thus are characterized by a specific collocational range. In terms of idiomaticity, collocations are said to be semantically more transparent than idioms, yet more opaque than free lexical combinations. The most important criterion for identifying collocation is the famous notion of mutual expectancy originally postulated by Firth (1957c). The constituents of a collocation mutually evoke each other, thus allowing a native speaker to predict the second partner when they encounter the first. So, two words that collocate are not governed by semantic compatibility but rather by lexical restriction, that is, by the norms of the language (see Coseriu, 1973). A consequence of this conventionalization is a certain associative bond that holds between the two partners of a collocation. Sinclair (1991: 119 ff.) highlights this function in postulating the idiom principle of language where he claims that for a large part of text production we use semi-preconstructed phrases that we choose at one go when speaking or writing.

We can conclude that collocations are conventionalized recurring word combinations exhibiting more or less restrictedness, more or less semantic opacity and a certain degree of predictability for native speakers. Their main function – besides disambiguating the meaning of words – is to facilitate smooth communication by reducing the processing effort for speakers and hearers alike (Wray, 2002). This fluency and