Collocations in corpora and in speakers' minds

Ewa Dąbrowska
Northumbria University

Although formulaic language has been studied extensively from both a cognitive linguistic and psycholinguistic perspective, the two approaches have developed relatively independently of each other. While most of the cognitive research has used corpus-based methods and has concentrated on identifying and describing formulaic units, most psycholinguists have focussed on the extent to which formulaic language facilitates processing, and more recently, on the psychological reality of the measures developed by corpus researchers. Relatively little is known about the relationship between an individual's knowledge of collocations and their linguistic experience, or between collocational knowledge and other aspects of linguistic knowledge, largely because research in these areas was hampered by lack of an adequate instrument measuring speakers' knowledge of collocations. This paper will describe the development of such an instrument and some preliminary research using it.

Part 1 of the paper describes the instrument itself. "Words that go together well" (Dąbrowska 2012a) is a multiple choice test consisting of 40 items. Each item consists of 5 short phrases: the target collocation (e.g. *raise prices*) and four foils (e.g. *elevate prices, grow prices, lift prices, stimulate prices*); participants are asked to select one that "sounds the most natural or familiar", and to guess when they were not sure. A preliminary list of collocations was extracted from a dictionary (Douglas-Kozłowska and Dzierżanowska 2004) and their collocational status validated using data from the British National Corpus. After initial piloting, the following criteria were adopted: overall frequency of at least 5 in the BNC **and** a mutual information (MI) score of at least 4. The foils had an MI of less than 2 and were not listed in the dictionary of collocations; the majority were also unattested in the corpus. The target items varied in frequency (from 5 to 260) and in collocation strength (MI from 4.2 to 13.1) and were selected in such a way that the two measures did not correlate to facilitate comparisons of their psychological relevance.

Part 2 describes a validation study of the test. Since knowledge of collocations is a function of the diversity of linguistic experience, it was predicted that scores should be positively correlated with measures of print exposure, education, and age, but not with non-verbal IQ. All four predictions were confirmed, although the correlation with age was significant for younger speakers only (below age 30).

Part 3 examines the relationship between speakers' knowledge of collocations and other aspects of linguistic knowledge, specifically, vocabulary size (assessed using the Vocabulary Size Test – Nation and Beglar 2007) and grammatical comprehension (assessed using the Pictures and Sentences Test – Dąbrowska 2012b). Usage-based constructionist approaches claim that all three aspects of speakers' knowledge are represented by means of the same type of declarative data structure (form-meaning pairings), and hence predict that all three should be correlated. These predictions were confirmed: all three correlations were moderately strong and statistically significant.

References

Dąbrowska, E. (2012a). Words that go together well. Available from
http://www.northumbria.ac.uk/static/5007/sasspdf/wordsthatgotogetherwell
Dąbrowska, E. (2012b). Pictures and Sentences. Available from
http://www.northumbria.ac.uk/static/5007/sasspdf/picturesandsentencesdescription
Douglas-Kozłowska, C., & Dzierżanowska, H. (2004). *Selected English Collocations*. Warszawa: PWN.
Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.