I don't know – differences in patterns of collocation and semantic prosody in phrases of different lengths

Bettina Starcke
University of Trier

Abstract

Different realisations of a lemma have different collocations (cf. Sinclair 1991 and Sinclair in Moon 1987: 89). This is also true for the different longer variants of a short phrase.

Taking the most frequent 3-gram in the BNC, 'i do n't', as a basis or nucleus phrase for the analysis, 4- to 6-grams which include 'i do n't' are analysed for their collocations and semantic prosodies. The results reveal that there are distinct differences in the usage of the n-grams. While the 3-gram 'i do n't' and the 4-gram 'i do n't know' collocate with hedging expressions and markers of uncertainty, their literal meaning, not knowing something, frequently disappears as their core meaning. In contrast to that, the 5-gram 'i do n't know what' and the 6-gram 'i do n't know what you' are mostly used in their literal sense which is negating knowledge of something. Unlike the 3- and the 4-grams, the longer phrases also have a distinct semantic prosody, namely that of anger, aggression, despair and frustration. A second, briefer study of phrases containing 'the end of' toward the end of the article will support the hypothesis that phrases of different lengths but with a shared nucleus phrase have different collocational patterns and distinct semantic prosodies.

In the second part of this article, explanations for these differences in collocations and semantic prosodies are offered. It is suggested that the longer a phrase, the more predetermined is its use. This is because semantic, pragmatic and syntactical restraints increase with the length of a phrase. Finally, implications of these findings for corpus linguistics in general are discussed.

1. Introduction

Language is frequently understood as a system of words which are formed into strings on the basis of grammatical rules. Words form into phrases, sentences, paragraphs and texts. The main unit of meaning and therefore also linguistic analysis has traditionally been the word. Only fairly recently have phrases become a focus in the analysis of language when linguists have recognised that they too carry meaning in language.

Since this finding has been relatively recent, basic concepts of the analysis of language have so far mainly been applied to the study of words and not to the study of phrases and phrasal meaning. This is for example the case with the concepts of collocation and semantic prosody.
1.1 Collocation and phraseology

The concept of collocation is one of the most essential in corpus linguistics. Firth discusses it as early as 1951 and later on describes it by saying that "you shall know a word by the company it keeps" (1958: 179). Leech (1974: 20) takes up the idea and introduces the concept of collocative meaning [which] consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment. (...) collocative meaning is simply an idiosyncratic property of individual words.

He then goes on to classify collocative meaning as an associative meaning and defines it as "What is communicated through association with words which tend to occur in the environment of another word" (26). Since it is an associative meaning, it is individual to the speakers of a language depending on whether they recognise this association or not.

The introduction of corpus linguistic techniques in the analysis of language has changed this perception of collocation as individual to the different speakers back to Firth's original notion. Collocation is now defined as the habitual co-occurrence of words as observable in a corpus. It is not intuitive anymore and the concept is firmly based on empirical evidence that was not available to Firth and Leech at the time when they wrote their definitions.

But not only are collocations intersubjective, empirical research has also revealed that different realisations of one lemma have different collocations. Sinclair (1991) first demonstrated that with an analysis of the different forms of YIELD in his discussion of how the lemma should be represented in a dictionary. What looked like a fairly easy concept – a lemma – turned out to be highly complex.

Partington (2004) confirms that different forms of one lemma, in his case study HAPPEN, occur in different contexts. While happens occurs in predominantly neutral contexts in his corpus of academic writings, happened mostly occurs in a negative context, and happen occurs twice as much in negative or neutral contexts as in positive ones. He also finds that this, as he calls it by referring to Hoey (2004), semantic priming is "realised within and through separate and typical phraseologies, characteristic syntactic patterns" (141).

While the differences in collocational patterning between different forms of lemmata have been discussed in depth, the application of this finding to phrases of various lengths has not been analysed. Phraseological research has frequently focussed on multi-word units as units of meaning (for example Sinclair 1996 and Stubbs 2005). Stubbs (2005) approaches this question by asking why the node world is among the most frequent words in the BNC. He finds that it is often part of a longer phrase such as best in the world which is extremely frequent in the language. It is therefore not the single word but the phrasal construction which is frequent. Consequently, phrases are seen to be the units of meaning in language and not single words. Words cannot be seen in isolation but have to be looked at in their contexts.