

A phraseology-driven approach to identifying evaluation in a book review corpus

Ute Römer

University of Hanover – Germany

Keywords: collocations, evaluation, book reviews, lexical-grammatical patterns, local grammar

1. Introduction

This paper presents selected results from a large-scale empirical study of evaluation in the language of English linguistic book reviews. In this study, evaluation is provisionally defined in a very broad sense, largely in line with Thompson and Hunston's (2000: 5) definition, as the expression of what a speaker or writer thinks of what s/he talks or writes about. This assessment of things that happen in the world can be positive or negative (see also Hunston 2004), or it can refer to the truth or likeliness of a proposition.

2. Towards a systematic analysis of evaluation in language

The study aims at a systematic rather than an eclectic account of evaluation, i.e. it tries to capture, as thoroughly as possible, in what ways evaluative meanings are created and conveyed in academic discourse, or more specifically in linguistic book reviews. This implies that the preferred evaluation detection strategy cannot just be to simply read a couple of reviews, compile a list of evaluative language items, and analyse the distribution and use of a small set of, let's say, negative evaluative adjectives (as done in Römer 2003).

Instead, methods have to be employed which provide insights on a higher level and uncover a larger number of linguistic means, the function of which is to evaluate. The method applied and described in the present paper is essentially phraseology-driven (referring back to Cowie's general definition of phraseology as "the study of the structure, meaning, and use of word-combinations", 1998: 673), meaning that it uses recurring word combinations (n-grams and collocations) to detect evaluative sequences in a review corpus.

3. Collocation extraction from BRILC

The corpus compiled for the analysis is BRILC, the Book Reviews In Linguistics Corpus, a monitor corpus of reviews published on the Web in Linguist List issues (between 1993 and 2005). BRILC currently consists of 1,310 individual book reviews and has a size of more than three million words. To access the corpus and retrieve collocations and n-gram lists, I am using the software program Collocate (Barlow 2004).

With respect to tracing evaluation, an extraction of the most frequent trigrams, 4-grams, and 5-grams in BRILC has led to a number of interesting observations. Topping the frequency-sorted 4-gram list, for instance, we find the combinations 'on the other hand', 'at the same time', 'the fact that the', and 'it would have been'. An analysis of the BRILC concordances of these and other common 4-grams shows that they either express evaluation themselves or appear in highly evaluative contexts. A few illustrative examples from the corpus are given in (1) to (8).

- (1) *On the other hand*, it is obvious that the book under review fails in various regards to take into account major developments in research into Indian English over the last 25 years.
- (2) *On the other hand*, all of my negative remarks are minor.
- (3) *At the same time*, I feel K does a good job of presenting the material in a form that should be accessible to readers who do not have a strong background in statistics.
- (4) [...] appear more radical and controversial than is really the case, *at the same time* not making it entirely clear how their own, alternative ideas would work in practice.
- (5) Another weak aspect is *the fact that the* analyses are based on official written transcripts that lack the natural spokenness of parliamentary talk.
- (6) Despite *the fact that the* book seemed to lack empirical evidence in some places, I could find little to argue with.
- (7) *It would have been* helpful if the author had been more explicit about what he regards as "a situation like this", i.e. under what circumstances the adposition is considered devoid of any meaning or function.
- (8) [...] most of the comparisons of the lexical sets are clearly useful and could have been extended, e.g. *it would have been* interesting to see the Australian vowels compared not just to RP, but also to American English vowels.

4. Lexical-grammatical patterns and local grammars

My approach is to some extent inspired by the local grammar approach in the Birmingham corpus linguistics tradition, in particular by Hunston and Sinclair's (2000) work on a local grammar of evaluation. Such a local grammar is basically a model "to deal with different forms of text organisation which occur with otherwise normal text" (Barnbrook 2002: 94). I argue that book reviews are organised in a particular way and would suggest seeing the language of evaluation that they contain as a specialised language with its own underlying system. This specialised language exhibits particular lexical-grammatical patterns – patterns of evaluation – many of which can be identified by means of the above-mentioned phraseology-driven method.

To pick out just one example, the 4-gram 'it would have been' only occurs in particular contexts in BRILC, forms certain patterns, and allows for a specific kind of variation. We find that *would* is sometimes replaced by the (so-called 'past tense') modals *could*, *might*, and *should*, which leads to the collocations 'it could have been', 'it might have been', and 'it should have been'. Also, we see in the concordance of 'it * have been' (with * being a wildcard character which stands for any one word, in this case a modal) that in most cases an adjective appears immediately to the right of the search string, as in examples (7) and (8) above.

