

Frequency analysis of English vocabulary and grammar

Stig Johansson and Knut Hofland

Oxford: Clarendon Press, 1989. Vol. 1, 400pp. Vol. 2, 380pp.

These books have been derived from a million-word collection of contemporary British English printed texts, known as the Lancaster–Oslo/Bergen Corpus — or LOB, for short. This corpus has been analysed at four levels: (a) frequency of word classes, (b) frequency of words, (c) combinations of word classes, and (d) combinations of words. Volume 1 gives the analysis of (a) and (b); Volume 2 of (c) and (d). Under (a), detailed information is given on the frequency of nouns, verbs, prepositions, etc. in the corpus as a whole and in the different text categories of the corpus (e.g. press, religion, science fiction). Under (b), we are given an alphabetical list of all the words in the corpus, including the frequency of different grammatical uses. Under (c), information is provided on which word classes immediately precede and follow any given word class. Under (d), combinations are given for a selection of frequent nouns, verbs, and adjectives.

The books are a major step forward in the provision of quantitative information about British English. For example, a traditional word-count of this corpus would tell you that *round* is used 335 times in a million words — useless information for a therapist or teacher, because you have no idea which parts of speech are involved. The present book gives you separate figures for *round* in all its uses — preposition, adjective, common noun, adverbial particle, and verb. Or again, under word combinations, you are told the number of occasions when *ask* was followed by *about*, *at*, *by*, and other forms (within a span of seven words). If you are devising teaching materials on a specific word or word class, and you want to make your sentences reflect closely the normal patterns of the language, you will need to know which is the most frequent adjacent word or word class. These books will be an invaluable aid to people involved in the creation of grammatical and lexical teaching materials. Once you have got used to the grammatical codes (conveniently summarised in an appendix), you use these books like you would a dictionary. And, like a dictionary, every resource centre should have one.