

Geoffrey Leech, Paul Rayson and Andrew Wilson. *Word Frequencies in Written and Spoken English*. 2001, xv + 304 pp. ISBN: 0 582 32007 0. Harlow: Pearson Education. Price: £45.99

Geoffrey Leech, Emeritus Professor in the Department of Linguistics and Modern English Language at Lancaster University, has been the co-editor and co-author of much research on English grammar, and computational and corpus linguistics. During the past thirty years, his major focus has been the analysis and processing of English with the building of the 100-million word British National Corpus, composed of modern English written texts as well as spoken transcriptions. Derived from a new version of this large electronic corpus, Leech *et al's* word-frequency book is a very comprehensive database for the linguist and lexicographer alike. In the past, frequency lists tended to focus on written language only, but this book has the merit of covering not only present-day written English, but also of encapsulating lexical frequency in modern spoken language. Spoken data only makes up 10% of the British National Corpus, however, as the transcription and digitisation of spoken language is extremely time-consuming. The spoken data in this book has been divided into a conversational part which includes informal dialogues (recordings of everyday spontaneous interaction) and a task-oriented part that represents 60% of the spoken data of the corpus and includes more formal audio material (lectures, TV and radio broadcasting).

The book begins with a 20-page introduction, which is sometimes quite technical, but which is indispensable for familiarising oneself with the data. It also gives guidelines on how to read the frequency lists. This introduction further discusses problems of data collection, transcription and spelling conventions. Each frequency list appears with a frequency score of tokens per million words. The book is structured into six main chapters that collect together the different word-frequency lists. Chapter 1 presents the whole corpus. The words are ranked without any distinction between speech and writing. Chapter 2 looks at word frequency across speech and writing. In chapter 3, the focus is on the spoken part of the BNC corpus with comparison between the conversational and task-oriented subcorpora. The written data of the corpus is the subject of chapter 4, comparing imaginative and informative writing. Thanks to a meticulous word-class tagging of nouns, adjectives, verbs etc., this book is not limited to word forms alone, but classifies lexical items according to their part of speech. In chapter 5, frequency tables give the most frequent words within each grammatical word class and the frequency of each grammatical class is presented in the final chapter.

The corpus used is varied enough, drawing from multiple sources, and is far more up-to-date than previous word lists representing tendencies of spoken as well as contemporary written English. For the lexicographer, this is an essential tool which throws light on the trends in modern English: it foregrounds words that belong to the core vocabulary of the language such as tool-

words (articles, prepositions, conjunctions etc.), modal verbs and auxiliaries, verbs of movement and thinking, and draws a contrast with words of lower frequency that would be worth including in a compact dictionary or a lexicographic collection, or on the contrary should be excluded in favour of more frequently-used forms or variants. The frequency lists allow the lexicographer to view the data sorted and presented according to different criteria (alphabetically, by descending order of frequency and by construction). The corpus is large enough for him/her to analyse trends and track changes diachronically or synchronically, establish the vitality of rare words, and is also available on the Internet in electronic form. For corpus linguists, this book is of paramount importance for analysing and establishing the saliency of lexical items with a comparative corpus, such as journalistic or scientific English.

Similar word-frequency lists have been compiled for Romance languages such as Spanish, Rumanian and French (Juilland *et al* 1964, 1965, 1970, Baudot 1992), but these studies all need updating. There also exist word lists in each respective language that have been devised using concordancers, but they are usually quite limited. It is hoped that Leech *et al*'s book will lead the way for the compilation of comprehensive word lists in other foreign languages.

References

- Baudot, J.** 1992. *Fréquences d'utilisation des mots en français écrit contemporain*. Montreal: Les Presses Universitaires de Montréal.
- Juilland, A., D. Brodin and C. Davidovitch.** 1970. *Frequency Dictionary of French Words*. The Hague/Paris: Mouton.
- Juilland, A. and E. Chang-Rodrigues.** 1964. *Frequency Dictionary of Spanish Words*. London/The Hague/Paris: Mouton.
- Juilland, A., P.P. Edwards and I. Juilland.** 1965. *Frequency Dictionary of Rumanian Words*. London/The Hague/Paris: Mouton.

Michaël Abecassis
Department of Modern Languages
University of Oxford
Oxford
United Kingdom
(*michael.abecassis@modern-languages.oxford.ac.uk*)