

Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture

Michael Stubbs
(University of Trier)

Blackwell Publishers (Language in Society series, edited by Peter Trudgill, volume 23), 1996, xix+267 pp; hardbound, ISBN 0-631-19511-4, \$59.95; paperbound, ISBN 0-631-19512-2, \$24.95

Reviewed by
Edward Finegan
University of Southern California

Text and Corpus Analysis appears in Blackwell's Language in Society series and is probably better characterized by its subtitle (*Computer-Assisted Studies of Language and Culture*) than by its title. Engaging in critical linguistics, Stubbs analyzes several independent texts and makes comparisons with some findings of well-known corpora. Critical linguistics is the enterprise whereby the politics and social values implicit in a text are uncovered by close linguistic analysis and made explicit. The central aim of the book under review is to demonstrate that some very simple tools of corpus linguistics can be enlisted in the analysis of ideology in language.

A major premise of the book and of critical linguistics more generally is this: "Grammar is choice. Different choices produce different meanings. The political nature of such choices is evident when analysis is applied to texts which represent relations between groups of people" (p. 98). In other words, "all linguistic usage encodes representations of the world. It is always possible to talk about the same thing in different ways, and the systematic usage of different syntactic patterns encodes different points of view" (p. 130). Stubbs lays out a basis for that premise by sketching the neo-Firthian view of language as articulated particularly in the sociolinguistics work of M.A.K. Halliday and the corpus linguistics work of John Sinclair. Stubbs also relates his analysis to a broader sociological analysis through the work of Anthony Giddens.

To exemplify critical linguistics, Stubbs analyzes five main texts that range in size from a few hundred to tens of thousands of words (the former as in Baden-Powell's farewell letters to the Girl Guides and Boy Scouts and the latter as in two secondary-school textbooks, the larger of which contains 80,000 words). For comparisons with his findings in these texts, as well as for pertinent examples of support and contrast throughout the book, and for exploring significant collocational patterns, he exploits particularly the one-million-word LOB and a 130-million-word British subcorpus of The Bank of English.

The principal computational tool that Stubbs uses in analyzing texts and their cultural and social values is the simple concordance. He exploits it, for example, to identify collocational patterns for what Raymond Williams calls *keywords* and Firth called *focal* or *pivotal* words—e.g., *standards*, *discipline*, *fundamentals*, *basics*, *common sense*, *literary heritage*, *nation*, and *proper education* (p. 159). The conclusion Stubbs draws from his analyses: "fixed and semi-fixed expressions (collocations, catch phrases, clichés and idioms) encode cultural information. Such recurrent phrases, which derive partly from works of literature and cultural history, are one of the ways in which experience is

represented and transmitted. The study of recurrent wordings is therefore of central importance in the study of language and ideology, and can provide empirical evidence of how the culture is expressed in lexical patterns" (p. 169).

At various points in *Text and Corpus Analysis*, Stubbs stresses that the principal object of linguistic analysis ought to be texts rather than sentences; this principle is important to him in the same way that he insists on the importance of naturally occurring linguistic data rather than intuitions. But his tools of analysis lead him at times to focus on sentences rather than whole texts, and his analyses sometimes suggest that he has equated a text with the sum of its sentences. For example, he draws inferences about the assignment of agency and responsibility by calculating frequency of passive voice verbs and nominalizations. In one analysis of a newspaper text reporting violence in South Africa, Stubbs concludes: "The violence is made more abstract and less brutal by the grammatical choices which are made. There is a connection between the grammar used and the perspectives adopted" (p. 96). Elsewhere he calculates frequencies of passive verbs, transitive and intransitive verbs, and inanimate subjects in order to gauge the degree of unexpressed agency in texts. Such equations may not be as straightforward as the analysis occasionally seems to suggest, however. While the equation may be justified in the analysis of a particular feature in a particular text, it is not justified in principle. Nor will selective exemplification of the equation be sufficient to demonstrate its consistent or systematic interpretation. Moreover, the Whorfian overtones of such analysis need special handling if they are not to be dismissed as impossible to disprove.

I would not quarrel with the basic premises of critical linguistics or the likelihood that Stubbs's methods of analysis of texts will often be perceptive and accurate. Rather, I would note, for example, that the interaction of active and passive voice verbs can also serve to create coherence in a text and reflect the application of sound principles of information flow. To examine texts without acknowledging multiple functions for particular linguistic features and to give the appearance of examining a set of sentences rather than a coherent text is to risk a charge of tunnel vision. With its melding of linguistics, literary analysis, and sociology, critical linguistics is in a position to help foster more sophisticated approaches to language analysis than are current in schools and the press, but that achievement will be frustrated if the enterprise can be brushed aside as myopic.

Text and Corpus Analysis may have limited appeal for those readers of this journal who are interested primarily in techniques of computational linguistics. Indeed, the few computational tools used by Stubbs are straightforward and widely available, and some chapters report no computational findings but simply rely on corpora to provide attested examples of the linguistic phenomena discussed. Rather, the book excels at illustrating how students and scholars might tackle the analysis of cultural, sociological, and political values in texts, using techniques of grammatical analysis and computational linguistics. Computational linguists should be aware of this book and refer students of sociology, politics, literature, and culture to it for exemplification of just what a computational linguistics approach can and cannot reveal. It should be noted that in this particular case, as in critical linguistics generally, the exemplifications do not fall along a wide range of the political or social spectrum. Stubbs is not merely illustrating—he is also arguing—and his arguments are distinctly egalitarian and progressive.

Ed Finegan is Professor of Linguistics and Law at the University of Southern California. With Douglas Biber he has developed ARCHER (A Representative Corpus of Historical English Regis-

ters). His textbook *Language: Its Structure and Use* (Harcourt Brace) will appear in its third edition in 1998. Finegan's address is Department of Linguistics, University of Southern California, Los Angeles, CA 90089-1693; e-mail: finegan@rcf.usc.edu