

Book Review

Dimensions of Register Variation: A Cross-Linguistic Comparison

Douglas Biber

(Northern Arizona University)

Cambridge University Press, 1995,
xvii+428 pp; hardbound, ISBN
0-521-47331-4, £37.50

Reviewed by
Michael Stubbs
University of Trier

Many readers will be familiar with the quantitative and distributional techniques of register analysis presented by Biber (1988). The present book re-uses this earlier study on variation in British English [henceforth E], along with "a synthesis" (p. xv) of Ph.D. studies done with Biber at the University of Southern California: by Niko Besnier on Nukulaelae Tuvaluan [T], Mohamed Hared on Somali [S], and Kim Yong-Jin on Korean [K], plus other work, especially by Ed Finegan and Dwight Atkinson. What is new in this book is its cross-linguistic comparisons of register. Biber finds striking and "highly systematic similarities" in the patterns of linguistic variation across languages, and this leads him to propose "the possibility of cross-linguistic universals governing the patterns of discourse variation" (p. 359).

The basis for this claim is the large differences between the four languages in their genetic affiliations, cultural histories, uses in literacy, and so on. T is Austronesian, with fewer than 350 speakers, on a small Pacific atoll. S is Afroasiatic, with 5 million speakers. And K is Altaic, with 65 million speakers. Appendices provide grammatical sketches of K and S, but not of T. And chapter 3 provides socio-cultural background on the four languages, some of it fairly rough and ready: for example, it is not mentioned that K is spoken in two different countries with very different political systems, and—I suppose—different registers. And, as Biber admits, E consists of a highly variable set of geographical and social varieties.

In chapters 1, 2, 4, and 5, Biber usefully overviews much previous work on register variation and the methods of register analysis that he developed in his 1988 work. He illustrates in detail the problems of trying to compare registers on the basis of isolated linguistic features. And he argues for an alternative multidimensional approach, where each dimension consists of a set of features that have been found to co-occur frequently in texts. He summarizes methods of grammatically tagging corpora, calculating the frequency of occurrence of features, using factor analysis to identify the dimensions, and then attributing functional interpretations to the dimensions. The main data comprise: 1.7 million words of E, 153,000 words of T, 136,000 words of K, and 480,000 words of S.

Chapters 6, 7, and 8 present the main findings. First, Biber shows differences in register variation across the languages. For example: T has a restricted range of registers, all interpersonal in some sense, with only two written registers, personal letters and sermons, and no mass media; K and E have a similar range of spoken and written registers, but K has a system of honorifics that constitute a separate dimension with

no parallel in E, S, or T. Of course, we don't need computer-assisted tagging and factor analysis to tell us that K has honorifics. However, the main claim is that, given the major social and linguistic differences across the four languages, the cross-linguistic similarities are far stronger than the differences. For example, all four languages have multiple dimensions that distinguish between prototypical speaking (interpersonal: on-line production) and writing (informational: careful, edited production). Biber also documents a complementary argument that the diachronic evolution of written registers has been strikingly similar, over very different periods of time, in S and E.

The book has two main kinds of interest: descriptive and methodological. First, Biber presents new linguistic descriptions and striking new hypotheses. (In this short review, it is not possible to give examples of the detailed quantitative descriptions of linguistic features given for dimensions.) In earlier work by others, register comparisons have been largely restricted to individual features (or at most a few features regarded as distinctive), but have not been based on co-occurrence relations across a wide range of features. There are few studies of register variation in non-Indo-European languages, and few diachronic register studies. Biber also poses interesting Halliday-type hypotheses about linguistic evolution, especially on the role of literacy in extending the range of linguistic variation.

Second, the work is a very clear case where new, computer-assisted methods have given access to new data, and thus opened up a series of research topics that were previously inconceivable. Biber provides precise methods of comparing highly complex objects: languages vary internally (synchronic register variation), but registers also vary internally (some are narrowly clustered around a central mean, but others have a large standard deviation; they may evolve diachronically; and there are both subregisters and systematic differences within texts, e.g., between the beginnings, middles, and ends).

There are certainly unanswered questions in both these areas. For example: (1) I can find no specification of how accurate the tagging software was, just a note (p. 85) that tagging was checked manually. (2) Biber emphasizes the importance of sampling for register diversity. But even the English-language data are very incomplete in this respect: e.g., the very large written genre of business correspondence is missing in the standard corpora that he uses. (3) I assume from textual evidence that Biber has linguistic competence in S, but not in K or T. But I can find no discussion of the possible limitations of an analyst carrying out analyses for languages that she or he does not know. (Was this not a standard topic in American structuralism and its aftermath in the 1950s?)

Indeed, this last point raises the question of who might be competent to thoroughly evaluate the book: three of the languages are not widely known, and I guess no one in the world speaks all four languages discussed here (your reviewer certainly doesn't) and, in addition, has expertise in factor analysis! So this short review can provide only a preliminary and provisional evaluation of the book.

Given the book's topic, it seems fair to make a comment on its academic style. Biber writes very clearly, but highly repetitively, and this long book would have benefited from a severe editing. An advantage of written academic language as a register is that it can be more condensed than spoken language, since readers can go back over parts that seem unclear. As it was, I frequently read slightly different formulations in different places, puzzled over whether a subtly different point was being made, and usually concluded that it wasn't. As noted, much of the repetition is due to the book having been "synthesized" from previous published work by Biber himself and his colleagues.

Over the last ten years, Biber has shown the possibilities of combining new meth-

ods and new data, first to document in more detail than ever before the nature of register variation, and now to formulate new hypotheses. But his cautiously formulated claim that he may have found stylistic universals will take many years and much further work to evaluate fully.

Reference

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge University Press.

Michael Stubbs is Professor of English Linguistics, University of Trier, Germany. He was previously Professor of English, Institute of Education, University of London (1985–90), and Chair of the British Association for Applied Linguistics (1988–91). His most recent book is *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture* (Blackwell, 1996). Stubbs's address is: FB2 Anglistik, Universität Trier, D-54286 Trier, Germany; e-mail: stubbs@uni-trier.de