

Douglas Biber and the Flagstaff School of corpus-based research

An introduction

 **Viviana Cortes**
Eniko Csomay

 <https://doi.org/10.1075/scl.66.004cor>

Pages xv–xx of

Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber

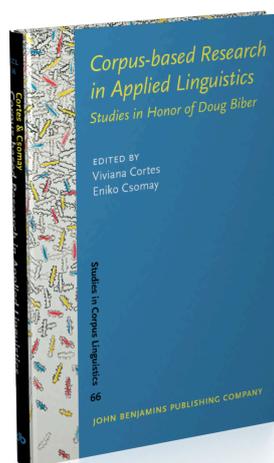
Edited by Viviana Cortes and Eniko Csomay

[*Studies in Corpus Linguistics*, 66] 2015. xix, 219 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights



INTRODUCTION

Douglas Biber and the Flagstaff School of corpus-based research

An introduction

Viviana Cortes & Eniko Csomay

Georgia State University / San Diego State University

The collection of essays in this book is a tribute to Professor Douglas Biber as our teacher and mentor, and in homage to the legacy of his teachings and research at Northern Arizona University (NAU, henceforth). As his disciples, we would like to call the space in which he taught and we learnt the tricks of the trade the Flagstaff School of Corpus-based Research. Undeniably, Prof. Biber and his work throughout his tenure at NAU have inspired many students, which resulted in a range of innovative corpus-based investigations.

Prof. Biber left his mark in each and every student who studied with him at NAU, inflicting his teachings with what we now recognize as the fundamentals of systematic and principled corpus linguistic research. These fundamentals are the pillars of the Flagstaff School, and they are the ones that distinguish it from any other existing or potential programs which are simply “doing” corpus-based research. In the Flagstaff School, we learnt how to:

1. Design and implement empirically-driven corpus-based research, paying close attention to the principles of Biber’s definition of corpus size, representativeness, sampling, and above all, to systematic analysis;
2. Actively engage in computer programming, allowing us not only to dare ask but to be able to answer corpus-based research questions never asked before. The reason these questions had not been asked before is that the tools available at the time did not allow the processing of texts to answer those questions. Instead of accepting that fact, however, we were challenged to design and create new tools of our own in order to satisfy our true curiosity and inquiry;

3. Place a strong emphasis on the combination of quantitative methods that are based on sound and innovative statistical procedures, and complemented with comprehensive qualitative functional analyses of language use.

Over the past decades, Biber's work has covered extensive grounds of linguistic inquiry, most prominently focusing on the study of register variation and corpus-based descriptions of grammar. His early work (Biber 1988) fundamentally challenged the traditional views of the dichotomy between speech and writing by providing empirical evidence through systematic descriptions of how, instead, spoken and written registers vary in a multidimensional linguistic space. This comprehensive description of register variation was applied in other languages as well, for example, Somali and Korean (Biber 1995) and most recently Spanish (Biber & Tracy-Ventura 2007), and in exploring the language of specific contexts, such as the university (Biber & Conrad 2009).

A decade after his first book in 1988, the *Longman Grammar of Spoken and Written English* (Biber et al. 1999) provided us with robust, corpus-based descriptions of English grammar never done before, and inspired many subsequent works. In addition, Biber has introduced the principles and techniques of corpus linguistics (Biber et al. 1998) and of corpus-based discourse analysis (Biber et al. 2007).

Each of the nine chapters selected for this volume was written following what he believed to be the main principles to do corpus research. The authors invited to write these chapters were, at some point in the past two decades, Prof. Biber's students in the Flagstaff School. We are now professors at different universities in the United States or other parts of the world, applying Biber's approach to corpus-based research and teachings to the current and future generations of corpus-based researchers. These authors have excelled in various areas of corpus-based research and their chapters represent each of those areas.

The first two chapters apply multi-dimensional approaches to the analysis of specific spoken registers. Eniko Csomay, in Chapter 1, investigates patterns of language use in presentations in the university setting. She focuses on two participants, teachers and students, as they present new information to an audience. She compiled a relatively small corpus of 168 teacher presentation segments in the classroom, and 76 student presentations recorded (and transcribed) at a student research symposium. Student and teacher presentations are then compared based on the dimensions of linguistic variation in university settings (Biber & Conrad 2009). Her findings show differences in language use between these two groups of presenters and these differences are attributed to the social status and the relationship between the speaker and their audience.

In Chapter 2, Eric Friginal uses multi-dimensional approaches to the description of spoken discourse, comparing telephone-based interactions in

three settings: (1) customer service transactions, (2) telephone conversations between friends, and (3) spontaneous telephone exchanges between participants discussing topics identified by fixed prompts. The findings indicate that variation in these interactions is largely influenced by the nature of conversational tasks, participants' roles in the interactions, and the use of the telephone as a medium in communicating ideas, opinions, and instructions.

The next four chapters analyze specific aspects of written discourse. Bethany Gray, in Chapter 3, uses a corpus of 270 (c. 2 million words) research articles as a single register and explores variation within that one register as it relates to the use of epistemic stance markers following the framework described in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999). The articles in her corpus showcase three distinctive research types (theoretical, qualitative, and quantitative) and are from six disciplines (Philosophy, History, Political Science, Applied Linguistics, Biology, and Physics). Tracking the lexico-grammatical features of stance with a special computer program she developed, the use of these stance markers are compared across disciplines and research article types.

In Chapter 4, Mohammed Albakry explores some of the linguistic and discursive aspects of framing positive and negative information in recommendation letters, using a corpus of 114 letters of recommendation to an English Ph.D. program. The findings show consistent patterns in the way different types of modals and their associated collocates are used to hedge predictions. In addition, through the analysis, the discursive frames of the most common mitigation strategies in presenting potentially negative information about applicants also become apparent.

In Chapter 5, Alfredo Urzua challenges the misconception that corpus linguistics relates to the de-contextualized nature of corpus data. To prove his point, he has designed and built a context-specific corpus of student writing produced by Spanish-speaking English learners (mostly freshman students from Mexico) at various levels of proficiency and reflecting a variety of writing tasks collected at the University of Texas El Paso. This corpus allows researchers to examine theoretical issues while helps educators to identify key pedagogical issues as they evaluate learners' needs in relation to practices and beliefs of the local culture. The chapter illustrates the various ways in which this corpus can be used to not only conduct empirical research on second language writing, but also to establish links to teaching, learning, and assessment.

Chapter 6 by Don Miller highlights the methodological challenges inherent in reliably capturing meaningful sets of vocabulary for instructional focus. An analysis of a 3.1 million-word corpus of introductory psychology textbooks suggests that, while comparatively large, and, thus, presumably representative of the lexical variability in the target domain, this corpus was unable to capture a stable list of "important" words. Findings highlight an important issue requiring further

investigation in corpus-based vocabulary research: the extent to which corpora – and the word lists based on them – reliably represent the lexical variability of their target domains.

Chandrika Balasubramanian takes a look at new varieties of English (New Englishes) in Chapter 7. More specifically, the study is an empirical investigation of spoken and written registers of contemporary Indian English. The first part of the paper outlines the theoretical bases for corpus construction for the study of international Englishes, and describes the corpus of 1.5 million words used in this study. The second part of the paper shows, through the investigation of two grammatical features (Wh-questions, and additive and restrictive circumstance adverbials) that an international English like Indian English shows the same kind of internal variation that the more traditional “native” varieties do.

In Chapters 8 and 9, Casey Keck and Viviana Cortes take a different direction, producing essays that review the state of the art in the study of two constructs that share a lot of features in common and are closely linked to the corpus-based research methodologies that originated with Biber’s work in the Flagstaff School. Keck, in Chapter 8, presents a chronological review of her own work on textual borrowing in the written production of non-native speaker university students. She emphasizes the use of tailor-made computer programs that facilitated the different stages of the research studies she conducted. Her chapter includes a careful description of the methodology used, the software designed, and the results of her analyses, as well as various implications of her findings for the teaching of academic writing to non-native speaker of English (NNSE) writers. In Chapter 9, Cortes writes about her area of specialization, lexical bundles, groups of three or more words that frequently recur in a register. She goes back in time to the origins of the use of corpus-driven methodologies to identify frequent formulaic expressions empirically rather than intuitively. The purpose of this chapter is to clearly describe the lexical bundle as a construct in the spectrum of formulaic language to avoid confusion in the method of identification and analysis of these expressions. Her chapter highlights the work of Biber et al. (1999) as a foundation for the many studies of lexical bundles that have been conducted in the past decade.

Finally, we would like to acknowledge our colleagues, who took some time from their very busy agendas to write and review for our volume. First, we would like to thank Michael McCarthy. When we first envisioned this volume, more than ten years ago, we asked Michael if he would write a preface to the volume highlighting Professor Biber’s contribution to the development of corpus-based research. He agreed then and he remembered that when we contacted him again a couple of years ago with the prospectus of this collection. Michael has always been a friend of the Flagstaff School, visiting the institution on several occasions sometimes with his students. Then we want to thank a group of scholars

who worked in the reviewing and helped in the editing process (in alphabetical order) Tony Berber-Sardinha, Scott Crossley, Stephanie Lindemann, David Oakey, Trevor Shankin, Heidi Vellenga, and Camila Vazquez. They all provided detailed and critical feedback to the writers that surely contributed to enrich the chapters presented in this volume.

References

- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: CUP.
DOI: 10.1017/CBO9780511621024
- Biber, Douglas. 1995. *Dimensions of Register Variation*. Cambridge: CUP.
DOI: 10.1017/CBO9780511519871
- Biber, Douglas & Conrad, Susan. 2009. *Register, Genre, and Style*. Cambridge: CUP.
DOI: 10.1017/CBO9780511814358
- Biber, Douglas & Tracy-Ventura, Nicole. 2007. Dimensions of register variation in Spanish. In *Working with Spanish Corpora*, Giovanni Parodi (ed.), 54–89. London: Continuum.
- Biber, Douglas, Conrad, Susan & Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP. DOI: 10.1017/CBO9780511804489
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas & Barbieri, Federica. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26: 263–86. DOI: 10.1016/j.esp.2006.08.003

