



HAL
open science

**Douglas Biber, Ulla Connor, Thomas A. Upton,
Discourse on the Move: Using Corpus Analysis to
Describe Discourse Structure**

Shirley Carter-Thomas

► **To cite this version:**

Shirley Carter-Thomas. Douglas Biber, Ulla Connor, Thomas A. Upton, Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure. *ASp - La revue du GERAS*, 2010, 57, pp.87-91. halshs-00664114

HAL Id: halshs-00664114

<https://shs.hal.science/halshs-00664114>

Submitted on 29 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pre-publication draft

Shirley Carter-Thomas

Télécom SudParis

& LATTICE CNRS (UMR 8094)

Douglas Biber, Ulla Connor and Thomas A. Upton

Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure

John Benjamins Publishing, 2007.

ISBN 978-90-272-2302-9

I was impatient to read this new book by Douglas Biber and colleagues and discover their views on using corpus-based approaches to investigate discourse structure. However, the first point to make about the book, volume 28 in the John Benjamins series *Studies in corpus linguistics*, is its rather particular authorial status. This is not a book written solely by the three authors named on the jacket; nor is it however an edited volume. It is instead a multi-authored work, and in addition to Douglas Biber, Ulla Connor and Thomas Upton, there are 6 further co-authors: Budsaba Kanoksilapatham, Molly Anthony, Kostyantyn Gladkov, Eniko Csomay, James K. Jones, and Casey Keck. There are two single-authored chapters, Chapter 4 by Kanoksilapatham and Chapter 8 by Csomay, and the other chapters are written by a combination of two to four of the nine authors (the reader has to return to the preface for the precise authors' names). As a result of this complex structure, the book seems to fall between two stools, containing perhaps neither the variety of a volume of edited articles, nor the more streamlined focussed organisation of a more traditionally authored book.

The first chapter provides a general overview of the methodological questions at issue. The 'lead' authors (Biber, Connor & Upton) begin by discussing the three definitions of discourse analysis identified by Schiffrin *et al* (2001): 1) the study of language in use, 2) the study of linguistic/discourse structure beyond the sentence, and 3) the study of social practices and ideological assumptions. They point out that - whereas until now corpus linguistic approaches have been primarily restricted to the first category, typically disregarding individual texts and higher level discourse structure - the aim of *Discourse on the Move* is to discuss the applications of corpus analysis to the analysis of discourse structure. Can corpora be used to reveal general patterns of discourse organisation, and can individual texts be analysed in terms of the general patterns that result from corpus analysis? The authors introduce the two major corpus-based approaches that can be applied to these research goals: i.e. top-down and bottom-up. In both, the segmentation of texts into analysable structural units is a major methodological focus. In top-down approaches the discourse unit types are determined before beginning the corpus analysis. Segmentation is based on identifying units of discourse that serve a particular communicative function and the qualitative-functional framework developed then extended to all texts in a genre-specific corpus. Linguistic-quantitative analysis comes at a later stage. In bottom-up approaches, the corpus analysis comes first. Texts are segmented according to quantifiable linguistic criteria and only after these units are already established is the qualitative description of their discourse functions performed. In both approaches, the analysis of the units identified makes use of multidimensional analysis, the statistical method developed by Biber (1988) to analyse linguistic variation. The main steps and lists of lexico-grammatical features identified by the Biber tagger are listed in the two appendices to the book.

Top-down methods for the analysis of discourse structure are presented in Chapters 2-5. Chapter 2, by Connor, Upton and Kanoksilapatham, focuses on the goals and methods of one particular type of top-down approach, that of genre-based move analysis. After presenting Swales' (1990, 2004) theory of move analysis as applied to research articles, the authors describe applications of moves analysis to other genres. They stress the importance of using corpus-based approaches within moves analysis in order to obtain generalisable descriptions of discourse organisation and discuss some of the important methodological issues involved such as coding protocols, move-tagging procedures, pilot-coding, interrater reliability and the use of statistical tests.

The next two chapters provide specific examples of corpus-based move analysis. In Chapter 3 (Upton and Connor), move analysis is applied to a 2 million word corpus of fundraising documents. The authors describe how fundraising letters are structured in terms of seven rhetorical moves and analyse the distribution and typical linguistic characteristics of stance structures within the moves identified. Chapter 4 is devoted to an in-depth study of biochemistry research articles. After identifying fifteen moves that articulate the four standard IMRD sections, Kanoksilapatham then uses Biber's multidimensional analysis to describe the typical linguistic characteristics of moves in her corpus. Forty-one linguistic features having strong patterns of variation are grouped into seven factor or dimension groupings and their distribution collated with the move types. The integration of moves analysis with a full-blown multi-dimensional analysis on a tagged corpus allows the author to pinpoint certain variational patterns which would arguably have gone unnoticed in approaches focusing solely on individual linguistic features.

Chapter 5 presents a second top-down approach to analysing discourse structure: appeals analysis. Exploiting the same corpus of fundraising letters as in Chapter 3, the authors (Connor, Anthony, Gladkov and Upton) analyse and segment the texts in terms of the three basic means of Aristotelian persuasion: rationality (logos), credibility (ethos), and affectivity (pathos). These three main appeals are subsequently subdivided into 19 appeal types. Rational appeals include, for example, the categories of authority, cause/effect and citing models. In a second stage, a linguistic descriptions of the appeals is undertaken, using various wordlists, keywords and concordance analyses. In the final section of the chapter the authors compare appeals analysis and moves analysis and point out the benefits of studying discourse structure from multiple perspectives. Certain appeals can be predicted to occur in certain move types and used together the two methods provide complementary information about the top-down structure of fundraising letters. All in all this chapter provides interesting reading but would perhaps have benefited from being placed immediately after the first analysis of fundraising letters, that is to say if the order of Chapter 4 and 5 had been reversed.

Bottom-up corpus based methods for analysing discourse structure are discussed in Chapters 6-8. The bottom-up approach advocated in this book relies on the identification of Vocabulary-Based Discourse Units (VBDUs), a largely automated approach inspired by Hearst's (1994, 1997) TextTiling programme. Chapter 6 (Biber, Csomay, Jones, and Keck) presents and discusses the method in detail. The TextTiling procedure basically analyses vocabulary patterns within the text, searching for sharp breaks in vocabulary and identifying the boundaries of units according to their similarity in vocabulary. Once the VBDUs are identified the segments are then analysed linguistically using different types of multidimensional analysis and VBDU text types are established. As the authors underline, linguistically founded text type distinctions do not necessarily correspond to move/step

distinctions (or even to typographical section breaks) and can thus provide different insights into text organisation tendencies.

Chapters 7 and 8 present two case-studies based on the bottom-up VDBU approach. In Chapter 7, Biber and Jones apply the method to a corpus of 100 biology research articles. After identifying the VBDUs and establishing the text types of each using multidimensional analysis, the authors interpret their results within the framework of the research articles' IMRD sections in order to identify other possible micro-genres operating within their scope. In Chapter 8, Csomay reports on a bottom-up VDBU analysis of oral academic discourse, that of university class sessions, from the T2K-Swal and MICASE corpora¹. Once again multidimensional analysis is used to analyse the resulting VBDUs. The features analysed include, in addition to the lexico-grammatical elements listed in the Bibber tagger, an analysis of 'turns': the number and average length of turns for teacher and students. Unlike research articles, university class sessions are not associated with a clearly defined detailed macrostructure. The ensuing qualitative analysis of the sequences of VBDUs allows the analysts to pinpoint certain variations in the class sessions that could not have been obtained had the analysis been based on the very simple top-down macrostructure (of only openings and closings).

In the final chapter, Biber, Connor and Upton discuss the relative merits of top-down and bottom-up approaches to discourse organisation. Contrasting the moves-based analysis of biochemistry articles performed in Chapter 4 with that of the VDBU segmented corpus of biology articles in Chapter 7, they conclude that both approaches to text segmentation produce discourse units that enable linguistic variation to be tracked successfully and can thus provide complementary insights into the discourse-organising patterns of a genre. The authors end the book by expressing the hope that the studies presented will inspire new research in the field of discourse organisation and suggest some possible future directions, ranging from the investigation of multimodal texts (Kress & Van Leuwen 2001) to the integration of ethnographic and contextual information, through surveys and interviews with informants (cf. Bazerman & Prior 2004; Connor 2000).

All in all, this book provides a useful synthesis of the issues involved in corpus-based analyses of discourse organisation. The methodological issues are clearly stated and the analytical procedure advocated is generally easy to follow. The most interesting contribution for me is undoubtedly the discussion on the integration of top-down approaches and corpus analysis, and the possibilities for extending such analyses to large-scale corpora discussed in the early chapters (particularly Chapters 2 and 3). However, I also feel that this book does not entirely live up to its promise. Although undoubtedly asking many of the right questions (see Chapter 1), no real new directions are suggested. Several of the analytical procedures presented have been around for many years (moves analysis, multidimensional analysis) and the text genres examined are perhaps rather too familiar (research articles, class discussion, fundraising letters). I regret too that only one type of bottom-up analysis was presented. Discourse analysts or computer linguists looking for new tools or methods may be disappointed.

References

¹ For a description of T2K-Swal corpus, see Biber *et al* 2004.
For the MICASE corpus, see <http://quod.lib.umich.edu/m/micase/2004>

- BAZERMAN, C. and P. Prior. 2004. *What writing does and how it does it: An introduction to analyzing texts and textual practices*. Mahwah, NJ: Lawrence Erlbaum.
- BIBER, D. 1988. *Variation across speech and writing*, Cambridge, Cambridge University Press.
- BIBER, D., S. CONRAD, R. REPPEN, P. Byrd, M. HELT, V. CLARK, V. CORTES, E. CSOMAY, and A. URZUA. 2004. *Representing language Use in the University: Analysis of the TOEFL 2000 Spoken & Written Academic Language Corpus (T2K-SWAL)*. TOEFL Monograph Series. Princeton, New Jersey : Educational Testing Service.
- CONNOR, U. 2000. Variation in rhetorical moves in grant proposals of US humanists and scientists, *Text*, 20(1), 1-28.
- HEARST, M. 1994. Multi-paragraph segmentation of expository texts. Technical Report 94/790, Computer Science Division (EECS), University of California, Berkeley.
- HEARST, M. 1997. TextTiling: segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.
- KRESS, G and VAN LEUWEN T. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold, 2001.
- SCHIFFRIN, D., D. TANNEN and H. HAMILTON. 2001. *The Handbook of Discourse Analysis*. Blackwell.
- SWALES, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, Cambridge.
- SWALES, J.M. 2004. *Research Genres*. Cambridge: Cambridge University Press.