1

Multi-Dimensional Analysis: A Historical Synopsis

Douglas Biber

Introduction

For many years, researchers have been interested in variation in language use across different situations: the description of "registers" and "register variation" (see, for example, Biber and Conrad 2009). Studies of registers have come from a wide range of disciplines and subdisciplines, including functional/sociolinguistics, applied linguistics, corpus/computational linguistics, composition/rhetoric studies, and communication research. (See the surveys of previous research in Biber and Atkinson 1994; Barbieri 2009.)

Many studies have described the situational and linguistic characteristics of a particular register, such as conversations, radio broadcasts, university lectures, grant proposals, PhD dissertations, and so on. However, the study of register can also be approached from a comparative perspective, investigating the patterns of variation among the range of registers in a discourse domain. Register variation is inherent in human language: a single speaker will make systematic choices in pronunciation, morphology, word choice, and grammar associated with different registers, reflecting the situational characteristics of those registers. The ubiquitous nature of register variation was noted early on by a number of scholars, for example:

> Each language community has its own system of registers . . . corresponding to the range of activities in which its members normally engage. (Ure 1982, 5)
>
> Register variation, in which language structure varies in accordance with the occasions of use, is all-pervasive in human language. (Ferguson 1983, 154)
>
> No human being talks the same way all the time. . . . At the very least, a variety of registers and styles is used and encountered. (Hymes 1984, 44)

However, despite its fundamental importance, there were few comprehensive linguistic analyses of register variation before the 1980s. This disregard was due mostly to methodological difficulties: until that time, it was simply not feasible to analyze the

full range of texts, registers, and linguistic characteristics required for a comprehensive analysis of register variation. However, with the availability of large online text corpora and computational analytical tools, such analyses became possible.

The Multi-Dimensional (MD) analytical approach was developed for this goal. Early MD studies investigated the relationships among spoken and written registers in English (e.g., Biber 1984, 1986b, 1988) followed by studies of spoken/written register variation in Nukulaelae Tuvaluan, Korean, and Somali (e.g., Biber and Hared 1994 on Somali; Kim and Biber 1994 on Korean). Biber (1995) summarizes these early MD studies and discusses crosslinguistic similarities and differences in the patterns of register variation.

In the following three decades, there have been numerous studies applying this framework to analyze the patterns of register variation in different languages, as well as the patterns of register variation within more specialized discourse domains of English. This chapter surveys these studies, with a focus on the kinds of research questions that have been addressed through MD Analysis.

Theoretical precursors to MD Analysis

The major innovation of the MD approach is that it provides a methodology to empirically analyze the ways in which linguistic features co-occur in texts and the ways in which registers vary with respect to those co-occurrence patterns. These goals are accomplished through computational analyses of a large corpus of texts, representing multiple registers, followed by a statistical analysis employing factor analysis.

The importance of linguistic co-occurrence had been noted in the 1970s by linguists such as Firth, Halliday, Ervin-Tripp, and Hymes. Brown and Fraser (1979, 38–39) observe that it can be "misleading to concentrate on specific, isolated [linguistic] markers without taking into account systematic variations which involve the co-occurrence of sets of markers." Ervin-Tripp (1972) and Hymes (1974) identify "speech styles" as varieties that are defined by a shared set of co-occurring linguistic features. Halliday (1988, 162) defines a register as "a cluster of associated features having a greater-than-random . . . tendency to co-occur."

However, despite these theoretical discussions, descriptions of registers and styles during this time were based on consideration of only a few linguistic features, with no empirical analysis of the co-occurrence relations among the linguistic characteristics. Several earlier sociolinguistic investigations claimed that registers vary along an underlying linguistic/functional parameter, and proposed a set of linguistic features associated with that parameter, thus giving at least implicit recognition to the importance of co-occurrence relations. Studies of this type include Ferguson (1959) on "high" versus "low" diglossic varieties; Bernstein (e.g., 1970) on restricted versus elaborated codes; Irvine (1979) on formal versus informal registers; and Ochs (1979) on planned versus unplanned discourse. A few other early researchers went further in proposing specific linguistic co-occurrence patterns associated with two parameters of variation. These include Chafe (1982), Chafe and Danielewicz (1987), and Longacre (1976). Chafe identifies two parameters—integration/fragmentation

and detachment/involvement—and posits a number of linguistic features associated with each parameter. Longacre also identifies two underlying parameters—projected time and temporal succession—and posits a group of features associated with each. These studies are important in that they recognize the need for analyses based on the co-occurrence relations in texts.

However, there are three major theoretical differences between these earlier investigations of register variation and the MD approach. First, apart from the Chafe and Longacre frameworks, most previous studies analyzed register variation in terms of a single underlying parameter, suggesting that there was a single basic situational distinction among registers. Second, most previous studies assumed that register variation could be analyzed in terms of simple, dichotomous distinctions, so that varieties are either formal or informal, planned or unplanned, and so on. And finally, none of these early approaches applied empirical methods to identify sets of co-occurring linguistic features. Rather, researchers proposed sets of features that seemed to work together, based on their perceptions and intuitions.

MD Analysis differs in all three respects. MD studies have demonstrated that no single parameter or dimension is adequate in itself to capture the full range of variation among registers in a language. Rather, different dimensions are realized by different sets of co-occurring linguistic features, reflecting different functional underpinnings (e.g., interactiveness, planning, informational focus). Second, the dimensions in MD studies are quantitative, continuous parameters of variation, which distinguish among a continuous range of texts or registers. For this reason, dimensions can be used to analyze the *extent* to which registers are similar (or different). And finally, sets of co-occurring linguistic features (which comprise the dimensions) are identified empirically using quantitative statistical techniques in the MD approach. In contrast, there is no guarantee that groupings of features proposed on intuitive grounds actually co-occur in texts, and in fact, subsequent MD analyses show that neither Longacre's parameters nor Chafe's parameters are completely accurate in identifying sets of linguistic features that actually co-occur regularly in English texts. In contrast, the statistical techniques used in MD studies provide a precise quantitative specification of the co-occurrence patterns among linguistic features in a corpus of texts.

There was, however, one precursor to MD Analysis that employed statistical analysis of linguistic co-occurrence patterns, and amazingly, that study was carried out more than twenty years before these other sociolinguistic investigations: John Carroll's (1960) study on "vectors of prose style." Carroll analyzed thirty-nine "objective" linguistic measures and twenty-nine "subjective" perceptual ratings. His corpus seems small by present-day standards: 150 prose text samples, consisting of about 300 words each. But when we consider the fact that Carroll apparently did all linguistic analyses by hand, the corpus is impressively large! Each linguistic feature was counted in each text sample, and then each text sample was perceptually rated for twenty-nine stylistic evaluations by eight judges. Carroll then applied a statistical factor analysis to reduce those variables to six underlying parameters of linguistic style. This study is remarkable in that it was carried out before the days of large computational corpora, automated tagging software, and computer-based statistical analysis packages. Although it is not framed as a study of register variation, Carroll's (1960) study can in many respects be regarded as the first MD investigation of linguistic variation among texts in English.

Early MD studies of English

The MD analytical framework was originally developed to investigate the linguistic differences between spoken and written discourse in English. Previous studies had come to a wide range of contradictory conclusions about the relationship between speech and writing, with some claiming fundamental differences between the two modes, while others claimed essentially no difference between the modes. MD Analysis was motivated by two major limitations in these previous studies: they were based on a restricted set of registers (often only a single register in speech and writing), and they typically investigated only a small set of linguistic characteristics. In contrast, MD Analysis was developed to investigate a wide range of spoken and written registers with respect to a comprehensive set of lexico-grammatical characteristics. The notion of linguistic co-occurrence was key as the organizing principle to describe how those linguistic characteristics worked together as a system defining the patterns of variation among spoken and written registers in English.

This line of research began with a PhD dissertation at the University of Southern California (Biber 1984), followed by early research articles on the methodology (referred to as "multi-feature / multi-dimensional analysis," Biber 1995) and the research findings regarding the relationships between spoken and written registers in English (Biber 1986a, b). Those studies compared registers in the Brown corpus (taken to represent written English) to registers in the London-Lund corpus (taken to represent spoken English). One restriction of these early studies is that the computer program for grammatical tagging was entirely rule-based with no lexicon. As a result, the analysis was based on analysis of only forty-two linguistic features. Five factors were extracted for the statistical analysis, but most of the linguistic interpretation focused on the first three dimensions: "Interactive vs edited text," "Abstract vs situated content," and "Reported vs immediate style."

Although its importance is not widely recognized, the computer program used for grammatical tagging provides the foundation for MD studies. Over the years, the grammatical tagger used for MD studies of English has gone through several major rounds of revision and extension, resulting in analysis of a much more comprehensive set of linguistic features. Thus, the 1984 tagger was rewritten in the mid-1980s to incorporate a lexicon of words together with their possible grammatical categories, and an extended set of rule-based algorithms. This revised tagger was used as the basis for the 1988 MD Analysis of English, based on analysis of sixty-seven linguistic features. Both of those taggers were written in PL/1 and ran on IBM mainframe computers. In the early 1990s, a new tagger was developed in Pascal to run on desktop computers. The current version has both probabilistic and rule-based components, and uses multiple large-scale dictionaries. The tagger achieves accuracy levels comparable to other existing taggers, but it is more robust than many taggers, with different probabilities for the processing of "oral" versus "literate" texts. The current tagger analyzes a larger set of linguistic characteristics than many other taggers (e.g., distinguishing simple past tense, perfect aspect, passive voice, and postnominal modifier for past participle verbs; identifying the gap position for *wh* relative clauses; identifying several different kinds of complement clause, the existence of *that*-complementizer deletion, and

several lexico-grammatical characteristics of the verbs/adjectives/nouns controlling complement clauses; etc.). As a result, recent MD analyses of English have been based on analysis of about 150 different linguistic features, compared to the 42 features included in the 1984 analysis.

Biber (1988) is the culminating study of the early phase of MD research. This book-length treatment focused on the research question of how speech and writing in English differed linguistically. The analysis identified five main dimensions of variation in a general corpus of spoken and written registers. The co-occurring groupings of features on each dimension reflect distinct functional relations, indicated by the interpretive labels for each one: *Involved versus informational production*, *Narrative discourse*, *Situation-dependent versus elaborated reference*, *Overt expression of persuasion*, and *Impersonal style*. Biber (1988, chapters 6–7), Biber (1995, chapters 5–7), and Conrad and Biber (2001, chapter 2) provide detailed descriptions of the statistical findings and justification for these interpretations (based on the shared communicative functions of the co-occurring linguistic features on each dimension plus the distribution of registers along each dimension).

Early MD studies of other languages

The first major extension to MD research was the application of this methodology to study spoken/written register variation in languages other than English. The central research question of these studies continued to be the linguistic relationship between speech and writing, asking whether the general patterns uncovered in the MD Analysis of English would hold true for other languages and cultures. Three studies were carried out in relatively short order following the early MD studies of English: Besnier's (1988) study of Nukulaelae Tuvaluan; Kim and Biber's (1994) study of Korean; and Biber and Hared's (1992, 1994) study of Somali.

Biber (1995) is a book-length treatment that summarizes the MD research findings from the four languages studied to that point. Those MD studies of register variation uncovered both surprising similarities and notable differences in the underlying dimensions of variation. From both theoretical and methodological perspectives, it is not surprising that each MD Analysis would uncover specialized dimensions that are peculiar to a given language. After all, each of these studies is based on the linguistic features found in that language, and the corpora analyzed include the distinctive registers found in that language. Given those differences, it was expected that the dimensions of variation that emerge for each different language would be fundamentally different, and this expectation was born out to some extent. For example, the MD Analysis of Somali identified a dimension interpreted as "Distanced, directive interaction," represented by optative clauses, first and second person pronouns, directional preverbal particles, and other case particles. Only one register is especially marked for the frequent use of these co-occurring features in Somali: personal letters. This dimension reflects the specialized inventory of grammatical devices in Somali combined with the particular communicative priorities of personal letters in Somali, which are typically interactive as well as explicitly directive.

However, the much more surprising finding to emerge from these early MD studies was the discovery of related dimensions of variation that occurred across all four languages. Two such dimensions are especially noteworthy: a dimension associated with "oral" versus "literate" discourse, and a dimension associated with narrative discourse. These dimensions are strikingly similar across languages in three major respects (cf. Biber 1995, especially chapter 7):

1. The co-occurring linguistic features that define the dimensions of variation in each language;
2. The functional correlates of the dimensions across languages; and
3. The linguistic/functional relations among analogous registers across languages.

Thus, these dimensions were proposed as candidates for universal parameters of register variation, leading to the study of several additional languages over the following years.

MD studies that have applied the 1988 model of variation to study specialized discourse domains

Completion of a "new" MD Analysis requires considerable work and technical expertise, both for developing the computer program for grammatical tagging and for carrying out the statistical factor analysis (see the chapter by Egbert and Staples in this volume). Many scholars have instead opted for an alternative approach: applying the dimensions identified and interpreted in Biber (1988) to study the patterns of variation among registers in more specialized discourse domains of English. Using the 1988 dimensions allows researchers to understand new registers or specialized sub-registers relative to the range of general spoken and written English registers analyzed in the 1988 study.

Because this approach is both efficient and highly informative, it has been widely applied to the study of specialized English registers over the last thirty years. For example, Biber (1987) applied the 1984 model of variation to compare the linguistic characteristics of British versus American written registers. Most studies of this type, however, have applied the 1988 model of variation; such studies began to appear shortly after publication of Biber (1988). For example, Biber and Finegan (1989) describe the historical evolution of English registers with respect to the 1988 dimensions (see also Biber and Finegan 1992, 1997). Similarly, Atkinson (1992) describes the historical evolution of medical research writing with respect to the 1988 dimensions (see also Atkinson and Royal Society [Great Britain] 1999). Biber (1991) studies the oral and literate characteristics of primary school reading materials. Biber and Finegan (1994) use the 1988 dimensions to compare the linguistic characteristics of the Introduction-Methods-Results-Discussion sections in present-day medical research articles, while Biber and Finegan (1994) use the 1988 dimensions to identify distinctive stylistic characteristics of famous eighteenth-century authors. Conrad and Biber (2001) is an edited compilation of MD studies, including several papers that use the 1988 model of variation published in the 1990s. These include Biber and Burges's study of gender differences in dramatic dialogue (Biber and Burges 2001), Helt's comparison of British versus American spoken registers

(Helt 2001), Conrad's comparison of textbooks and journal articles in biology and history (Conrad 2001), Connor-Linton's stylistic comparison of individual texts about American nuclear arms policy (Connor-Linton 2001), Reppen's study of variation among elementary student registers (Reppen 2001), and Rey's study of female-male differences in the dialogue of *Star Trek* episodes from 1966 to 1993 (Rey 2001).

Applications of this approach continue to appear up to the present day, including a comparison of spoken versus written university registers (Biber et al. 2002); comparisons of TV dialogue with natural conversation (Quaglio 2009); a description of student written registers in Black South African English (Van Rooy 2008); a description of the patterns of register variation in East African English (Van Rooy et al. 2010); an analysis of academic registers written by university students (Nesi and Gardner 2012); a comparison of internet and pre-internet registers (Berber Sardinha 2014); a description of learner language in oral interviews (Perez-Paredes and Sanchez-Tornel 2015); and a comparison of translated English and a non-native indigenized variety of English (Kruger and Van Rooy 2016).

Later MD studies of other languages

Following the early publication of MD studies on Tuvaluan, Somali, and Korean, there was a considerable break in MD research investigating register variation in other languages. During the late 1990s, dissertations were completed by Jang (1998) on Taiwanese and Sáiz (1999) on Spanish. But it was not until well after 2000 that we begin to find new published MD studies of languages other than English. These include studies of register variation in Spanish (Asencion-Delaney 2014; Asencion-Delaney and Collentine 2011; Biber et al. 2006; Parodi 2007), Brazilian Portuguese (Berber Sardinha et al. 2014a, b), Czech (Kodytek 2007), Scottish Gaelic (Lamb 2008), Dagbani (Purvis 2008), and written Chinese (Z.-S. Zhang 2012). These more recent MD studies of other languages provide additional support for the existence of universal dimensions associated with the oral/literate opposition and the narrative/non-narrative opposition (see the conclusion for discussion).

"New" MD studies of register variation in English

The most unanticipated development for MD research is the large number of "new" MD studies that have been carried out for English register variation after publication of the 1988 model. That is, the 1988 study was framed as a description of the overall patterns of register variation in English. It was believed at that time that this methodological approach was highly effective for the analysis of linguistic variation across the full set of spoken and written English registers, but that the approach would be less effective for analyses of variation in more restricted discourse domains. This skepticism was based on the expectation that there would be insufficient variation in a restricted discourse domain to achieve stable or interpretable results in a factor analysis. However, that expectation has repeatedly been proven to be incorrect. In fact, the development of

Table 1.1 MD studies of specialized discourse domains in English

Study	Discourse domain
Biber and Burges (2001)	Eighteenth-century speech-based and written registers
Egbert (2012)	Nineteenth-century fictional style
Friginal (2008, 2009)	Call-center discourse
Biber (2008)	Conversational registers
Gray (2013, 2015)	Disciplinary differences across research articles
Biber and Jones (2005)	Discourse unit types in biology research articles
Reppen (2001)	Elementary student registers
Gardner et al. (submitted)	L1 English student written registers
Hardy and Römer (2013)	L1 English student written registers
Biber and Gray (2013)	L2 English student written registers
Biber et al. (2016)	L2 English student written registers
Van Rooy and Terblanche (2009)	L2 English student written registers
Cao and Xiao (2013)	L1 vs. L2 English abstracts
Hardy and Friginal (2012)	L1 vs. L2 English online communication
Kanoksilapatham (2007)	Move types in biochemistry research articles
Veirano Pinto (2014)	Movies
Bertoli-Dutra (2014)	Pop songs
Egbert (2014, 2015)	Publication type and academic discipline
Grieve (2014)	Regional dialects
Biber (2006)	Spoken and written university registers
Kruger and Van Rooy (2016)	Translated written registers
Berber Sardinha (2018)	TV registers
Berber Sardinha (2018)	Web registers
Biber and Egbert (2016)	Web registers
Titak and Roberson (2013)	Web registers
Xiao (2009)	World Englishes
Gozda-Roszkowski (2011)	Written legal registers

new MD models for specialized discourse domains of English has been by far the most productive avenue of MD research over the last twenty years (Table 1.1).

Methodological extensions to MD Analysis

One early extension to MD Analysis was the study of "text types," which are defined as groupings of texts that are similar in their linguistic characteristics. These grouping of texts are identified statistically by applying a cluster analysis, with the dimensions of variation used as the linguistic predictors to measure the extent to which any two texts

Table 1.2 MD studies based on other linguistic characteristics

Study	Linguistic focus
Berber Sardinha (this volume)	Bigrams
Crossley and Louwerse (2007)	Bigrams
Crossley et al. (2014, this volume)	Cohmetrix measures
Berber Sardinha (2017a)	Collocations
Fitzsimmons-Doolan (2014)	Collocations
Xiao and McEnery (2005)	Keywords
Zhang (2016)	Metadiscourse markers
Berber Sardinha (2015)	Metaphor use

are linguistically similar. There can be a considerable range of situational variation among the texts belonging to a single text type, but all texts grouped into a "type" are linguistically similar. Early studies in this line of research analyzed the text types of English and Somali (Biber 1989, 1995, chapter 9, 2013). More recently, this approach has been applied to study web text types (Biber and Kurjian 2007) and text types in Brazilian Portuguese (Berber Sardinha 2017b).

Other studies have explored the extent to which similar linguistic generalizations can be achieved through alternative statistical techniques (i.e., other than exploratory factor analysis). Biber (1992, 2007) applies confirmatory factor analysis to identify the best-fitting MD model of complexity features. Several other studies have focused on the automatic classification of texts into register categories, and thus have coupled MD Analysis with predictive techniques like discriminant analysis (see Berber Sardinha and Veirano Pinto 2015; Biber 1993; Biber and Egbert 2016; de Mönnink et al. 2003; Veirano Pinto in this volume). Egbert and Biber (2016) directly compare the research findings achieved through a factor analysis to those achieved through a discriminant analysis, concluding that the two are in many ways compatible because they both reflect communicative functions.

Finally, several studies have explored the extent to which interpretable dimensions of variation can be discovered based on analysis of linguistic systems other than the set of lexico-grammatical features that are traditionally considered in an MD Analysis (Table 1.2).

Conclusion

As noted above, probably the most theoretically important finding to emerge from the body of MD studies is the existence of universal parameters of register variation. From both theoretical and methodological perspectives, it is to be expected that each MD Analysis would uncover specialized dimensions that are peculiar to a given language and/or discourse domain. After all, each of these studies differs with respect to the set of linguistic features included in the analysis, and the set of registers represented in the corpus for analysis. Given those differences, it is reasonable to expect that the

parameters of variation that emerge from each analysis will be fundamentally different. And to some extent, this expectation is met with specialized dimensions emerging in nearly all MD analyses.

However, given this background, the more important finding is the existence of dimensions of variation that emerge in nearly all MD studies. Two such dimensions are especially important, regardless of the language or discourse domain: a dimension associated with "oral" versus "literate" discourse, and a dimension associated with narrative discourse.

The robustness of narrative dimensions across languages and discourse domains indicates that this rhetorical mode is basic to human communication, whether in speech or in writing. Rhetoricians and discourse analysts have long argued for the central role of narration in communication. MD studies confirm that claim, showing the importance of this rhetorical mode in virtually all discourse domains (spoken and written, interpersonal and informational, etc.).

But the most surprising pattern discovered through MD Analysis is the oral/literate opposition, which emerges as the very first dimension in nearly all MD studies (cf. Biber 2009, 2014) (cf. Biber 2009, 2014). In studies based on general corpora of spoken and written registers, this dimension clearly distinguishes between speech and writing. However, other studies show that this is not a simple opposition between the spoken and written modes. In fact, this dimension has emerged consistently in studies restricted to only spoken registers, as well as studies restricted to registers in the written mode.

In terms of communicative purpose, the "oral" registers characterized by this dimension focus on personal concerns, interpersonal interactions, and the expression of stance. In contrast, "literate" registers focus on the presentation of propositional information, with little overt acknowledgment of the audience or the personal feelings of the speaker/writer. These registers are usually produced in situations that allow for extensive planning and even revising and editing of the discourse.

Linguistically, this oral/literate dimension opposes two discourse styles: an "oral" style that relies on pronouns, verbs, adverbs, and a "literate" style that relies on nouns and nominal modifiers. The oral style relies on clauses to construct discourse—including a dense use of dependent clauses. In contrast, the complexity of the literate style is phrasal. This finding, replicated across languages and across discourse domains, is especially surprising, because it runs counter to assumptions about syntactic complexity held by many linguists. But it is perhaps the most important and robust finding to emerge crosslinguistically from MD studies: spoken registers (and "oral" written registers) rely on clausal discourse styles, including a dense use of dependent clauses; written registers (and "literate" spoken registers) rely on phrasal discourse styles, especially the dense use of phrasal modifiers embedded in noun phrases (see also Biber and Gray 2011, 2016; Biber et al. 2011).

In sum, the patterns of variation observed across MD studies support the likelihood of universal parameters of register variation as well as the existence of unique dimensions of variation in each language and/or discourse domain. Future MD studies should further our understanding of both.

References

Asencion-Delaney, Y. (2014), "A Multi-Dimensional Analysis of Advanced Written L2 Spanish," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 241–75, Amsterdam/Philadelphia, PA: John Benjamins.

Asencion-Delaney, Y. and J. Collentine (2011), "A Multi-Dimensional Analysis of a Written L2 Spanish Corpus," *Applied Linguistics*, 32 (3): 299–322.

Atkinson, D. (1992), "The Evolution of Medical Research Writing from 1735 to 1985: The Case of the 'Edinburgh Medical Journal'," *Applied Linguistics*, 13: 337–74.

Atkinson, D. and Royal Society (Great Britain). (1999), *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675–1975*, Mahwah, NJ: L. Erlbaum Associates.

Barbieri, F. (2009), "Appendix A: Annotation of Major Register/Genre Studies," in D. Biber and S. Conrad (eds), *Register, Genre, and Style*, 271–95, Cambridge/New York: Cambridge University Press.

Berber Sardinha, T. (2014), "25 Years Later: Comparing Internet and Pre-Internet Registers," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 81–105, Amsterdam/Philadelphia, PA: John Benjamins.

Berber Sardinha, T. (2015), "Register Variation and Metaphor Use: A Multi-Dimensional Perspective," in J. B. Herrmann and T. Berber Sardinha (eds), *Metaphor in Specialist Discourse: Investigating Metaphor Use in Technical, Scientific and Popularized Discourse Contexts*, 17–52, Amsterdam/Philadelphia, PA: John Benjamins.

Berber Sardinha, T. (2017a), "Lexical Priming and Register Variation," in M. Pace-Sigge and K. Patterson (eds), *Lexical Priming: Applications and Advances*, 190–230, Amsterdam/Philadelphia, PA: John Benjamins.

Berber Sardinha, T. (2017b), "Text Types in Brazilian Portuguese: A Multidimensional Perspective," *Corpora*, 12 (3): 483–515.

Berber Sardinha, T. (2018), "Dimensions of Variation across Internet Registers," *International Journal of Corpus Linguistics*, 23 (2): 125–57.

Berber Sardinha, T., C. Kauffmann and C. Mayer Acunzo (2014a), "Dimensions of Register Variation in Brazilian Portuguese," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 35–80, Amsterdam/Philadelphia, PA: John Benjamins.

Berber Sardinha, T., C. Kauffmann and C. Mayer Acunzo (2014b), "A Multidimensional Analysis of Register Variation in Brazilian Portuguese," *Corpora*, 9 (2): 239–71.

Berber Sardinha, T. and M. Veirano Pinto (2015), "Predicting American Movie Genre Categories from Linguistic Characteristics," *Journal of Research Design and Statistics in Linguistics and Communication Science*, 2 (1): 75–102.

Bernstein, B. (1970), *Class, Code, and Control* (Vol. 1: Theoretical Studies toward a Sociology of Language), London: Routledge & Kegan Paul.

Bertoli-Dutra, P. (2014), "Multi-Dimensional Analysis of Pop Songs," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 149–75, Amsterdam/Philadelphia, PA: John Benjamins.

Besnier, N. (1988), "The Linguistic Relationships of Spoken and Written Nukulaelae Registers," *Language*, 64: 707–36.

Biber, D. (1984), "A Model of Textual Relations within the Written and Spoken Modes," PhD dissertation, University of Southern California.

Biber, D. (1986a), "On the Investigation of Spoken/Written Differences," *Studia Linguistica*, 40: 1–21.
Biber, D. (1986b), "Spoken and Written Textual Dimension in English: Resolving the Contradictory Findings," *Language*, 62: 384–414.
Biber, D. (1987), "A Comparison of British and American Writing," *American Speech*, 62: 99–119.
Biber, D. (1988), *Variation across Speech and Writing*, Cambridge: Cambridge University Press.
Biber, D. (1989), "A Typology of English Texts," *Linguistics*, 27: 3–43.
Biber, D. (1991), "Oral and Literate Characteristics of Selected Primary School Reading Materials," *Text*, 11 (73–96).
Biber, D. (1992), "On the Complexity of Discourse Complexity: A Multidimensional Analysis," *Discourse Processes*, 15: 133–63.
Biber, D. (1993), "Using Register-Diversified Corpora for General Language Studies," *Computational Linguistics*, 19: 219–41.
Biber, D. (1995), *Dimensions of Register Variation: A Cross-Linguistic Comparison*, Cambridge: Cambridge University Press.
Biber, D. (2006), *University Language: A Corpus-Based Study of Spoken and Written Registers*, Amsterdam/Philadelphia, PA: John Benjamins.
Biber, D. (2007), "On the Complexity of Discourse Complexity: A Multidimensional Analysis," in T. A. Van Dijk (ed), *Discourse Studies*, vol. 1, 127–57, London: Sage.
Biber, D. (2008), "Corpus-Based Analyses of Discourse: Dimensions of Variation in Conversation," in V. Bhatia, J. Flowerdew and R. Jones (eds), *Advances in Discourse Studies*, 110–14, London: Routledge.
Biber, D. (2009), "Multi-Dimensional Approaches," in A. Lüdeling and M. Kytö (eds), *Corpus Linguistics: An International Handbook*, vol. 2, 822–55, Berlin/New York: Walter de Gruyter.
Biber, D. (2013), "A Typology of English Texts (50-Year Jubilee Issue of the Journal)," *Linguistics*, 27 (1): 3–43.
Biber, D. (2014), "Using Multi-Dimensional Analysis to Explore Cross-Linguistic Universals of Register Variation," *Languages in Contrast*, 14 (1): 7–34.
Biber, D. and D. Atkinson (1994), "Register: A Review of Empirical Research," in D. Biber and E. Finegan (eds), *Sociolinguistic Perspectives on Register*, 351–85, Oxford: Oxford University Press.
Biber, D. and J. Burges (2001), "Historical Change in the Language Use of Women and Men: Gender Differences in Dramatic Dialogue," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 157–70, Harlow: Longman.
Biber, D. and S. Conrad (2009), *Register, Genre, and Style*, Cambridge/New York: Cambridge University Press.
Biber, D., S. Conrad, R. Reppen, P. Byrd and M. Helt (2002), "Speaking and Writing in the University: A Multi-Dimensional Comparison," *TESOL Quarterly*, 36 (1): 9–48.
Biber, D., M. Davies, J. K. Jones and N. Tracy-Ventura (2006), "Spoken and Written Register Variation in Spanish: A Multi-Dimensional Analysis," *Corpora*, 1 (1): 1–37.
Biber, D. and J. Egbert (2016), "Register Variation on the Searchable Web: A Multi-Dimensional Analysis," *Journal of English Linguistics*, 44 (2): 95–137.
Biber, D. and E. Finegan (1989), "Drift and the Evolution of English Style: A History of Three Genres," *Language*, 65: 487–517.
Biber, D. and E. Finegan (1992), "The Linguistic Evolution of Five Written and Speech-Based English Genres from the 17th to the 20th Centuries," in M. Rissanen et al. (eds),

History of Englishes: New Methods and Interpretations in Historical Linguistics, 688–704, Berlin: Mouton.

Biber, D. and E. Finegan (1994a), "Intra-Textual Variation within Medical Research Articles," in N. Oostdijk and P. de Haan (eds), *Corpus-Based Research into Language*, 201–22, Amsterdam: Rodopi.

Biber, D. and E. Finegan (1994b), "Multi-Dimensional Analyses of Authors' Styles: Some Case Studies from the Eighteenth Century," in D. Ross and D. Brink (eds), *Research in Humanities Computing 3*, 3–17, Oxford: Oxford University Press.

Biber, D. and E. Finegan (1997), "Diachronic Relations among Speech-Based and Written Registers in English," in T. Nevalainen and L. Kahlas-Tarkka (eds), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, 253–75, Helsinki: Societe Neophilologique.

Biber, D. and E. Finegan (2001), "Diachronic Relations among Speech-Based and Written Registers in English," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 66–83, Harlow: Longman.

Biber, D. and B. Gray (2011), "Grammar Emerging in the Noun Phrase: The Influence of Written Language Use," *English Language and Linguistics*, 15: 223–50.

Biber, D. and B. Gray (2013), "Identifying Multi-Dimensional Patterns of Variation across Registers," in M. Krug and J. Schlúter (eds), *Research Methods in Language Variation and Change*, 404–20, Cambridge: Cambridge University Press.

Biber, D. and B. Gray (2016), "Grammar Emerging in the Noun Phrase: The Influence of Written Language Use," *English Language and Linguistics*, 15 (20th Anniversary Collection): 223–50.

Biber, D., B. Gray and K. Poonpon (2011), "Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development?," *TESOL Quarterly*, 45: 5–35.

Biber, D., B. Gray and S. Staples (2016), "Predicting Patterns of Grammatical Complexity across Language Exam Task Types and Proficiency," *Applied Linguistics*, 37 (5): 639–68.

Biber, D. and M. Hared (1992), "Dimensions of Register Variation in Somali," *Language Variation and Change*, 4: 41–75.

Biber, D. and M. Hared (1994), "Linguistic Correlates of the Transition to Literacy in Somali: Language Adaptation in Six Press Registers," in D. Biber and E. Finegan (eds), *Sociolinguistic Perspectives on Register*, 182–216, Oxford: Oxford University Press.

Biber, D. and J. K. Jones (2005), "Merging Corpus Linguistic and Discourse Analytic Research Goals: Discourse Units in Biology Research Articles," *Corpus Linguistics and Linguistic Theory*, 1 (2): 151–82.

Biber, D. and J. Kurjian (2007), "Towards a Taxonomy of Web Registers and Text Types: A Multi-Dimensional Analysis," in M. Hundt, N. Nesselhauf and C. Biewer (eds), *Corpus Linguistics and the Web*, 109–32, Amsterdam/New York, NY: Rodopi.

Brown, P. and C. Fraser (1979), "Speech as a Marker of Situation," in K. Scherer and H. Giles (eds), *Social Markers in Speech*, 33–62, Cambridge: Cambridge University Press.

Cao, Y. and R. Xiao (2013), "A Multi-Dimensional Contrastive Study of English Abstracts by Native and Non-Native Writer," *Corpora*, 8: 209–34.

Carroll, J. B. (1960), "Vectors of Prose Style," in T. A. Sebeok (ed), *Style in Language*, 283–92, Cambridge: Technology Press of Massachusetts Institute of Technology.

Chafe, W. L. (1982), "Integration and Involvement in Speaking, Writing and Oral Literature," in D. Tannen (ed), *Spoken and Written Language: Exploring Orality and Literacy*, 35–54, Norwood, NJ: Ablex.

Chafe, W. L. and J. Danielewicz (1987), *Properties of Spoken and Written Language*, Cambridge, MA: Academic Press.
Connor-Linton, J. (2001), "Author's Style and Worldview: A Comparison of Texts about American Nuclear Arms Policy," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 84–93, Harlow: Longman.
Conrad, S. (2001), "Variation among Disciplinary Texts: A Comparison of Textbooks and Journal Articles in Biology and History," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 94–107, Harlow: Longman.
Conrad, S. and D. Biber (2001), *Variation in English: Multi-Dimensional Studies*, Harlow: Longman.
Crossley, S. and M. M. Louwerse (2007), "Multi-Dimensional Register Classification Using Bi-Grams," *International Journal of Corpus Linguistics*, 12 (4): 453–78.
Crossley, S., L. K. Varner and D. McNamara (2014), "A Multi-Dimensional Analysis of Essay Writing: What Linguistic Features Tell Us about Situational Parameters and the Effects of Language Functions on Judgments of Quality," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 344–411, Amsterdam/Philadelphia, PA: John Benjamins.
de Mönnink, I. M., N. Brom and N. H. J. Oostdijk (2003), "Using the MF/MD Method for Automatic Text Classification," in S. Granger and S. Petch Tyson (eds), *Extending the Scope of Corpus Based Research: New Applications New Challenges*, 15–25, Amsterdam: Rodopi.
Egbert, J. (2012), "Style in Nineteenth Century Fiction: A Multi-Dimensional Analysis," *Scientific Study of Literature*, 2 (2): 167–98.
Egbert, J. (2014), "Student Perceptions of Stylistic Variation in Introductory University Textbooks," *Linguistics and Education*, 25: 64–77.
Egbert, J. (2015), "Publication Type and Discipline Variation in Published Academic Writing: Investigating Statistical Interaction in Corpus Data," *International Journal of Corpus Linguistics*, 20: 1–29.
Egbert, J. and D. Biber (2016), "Do All Roads Lead to Rome?: Modeling Register Variation with Factor Analysis and Discriminant Analysis," *Corpus Linguistics and Linguistic Theory*, Ahead of print.
Ervin-Tripp, S. (1972), "On Sociolinguistic Rules: Alternation and Co-occurrence," in J. Gumperz and D. Hymes (eds), *Directions in Sociolinguistics*, 213–50, New York: Holt, Rinehart and Winston.
Ferguson, C. A. (1959), "Diglossia," *Word*, 15: 325–40.
Ferguson, C. A. (1983), "Sports Announcer Talk: Syntactic Aspects of Register Variation," *Language in Society*, 12: 153–72.
Fitzsimmons-Doolan, S. (2014), "Using Lexical Variables to Identify Language Ideologies in a Policy Corpus," *Corpora*, 9 (1): 57–82.
Friginal, E. (2008), "Linguistic Variation in the Discourse of Outsourced Call Centers," *Discourse Studies*, 10: 715–36.
Friginal, E. (2009), *The Language of Outsourced Call Centers: A Corpus-Based Study of Cross-Cultural Interaction*, Amsterdam/Philadelphia, PA: John Benjamins.
Gardner, S., D. Biber and H. Nesi (submitted). *Submitted to IJCL*.
Gozda-Roszkowski, S. (2011), *Patterns of Linguistic Variation in American Legal English: A Corpus-Based Study*, Frankfurt am Main: Peter Lang.
Gray, B. (2013), "More than Discipline: Uncovering Multi-Dimensional Patterns of Variation in Academic Research Articles," *Corpora*, 8: 153–81.
Gray, B. (2015), *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*, Amsterdam/Philadelphia, PA: John Benjamins.

Grieve, J. (2014), "A Multi-Dimensional Analysis of Regional Variation in American English," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 3–34, Amsterdam/Philadelphia, PA: John Benjamins.

Halliday, M. A. K. (1988), "On the Language of Physical Science," in M. Ghadessy (ed), *Registers of Written English*, 162–78, London: Pinter.

Hardy, J. and U. Römer (2013), "Revealing Disciplinary Variation in Student Writing: A Multi-Dimensional Analysis of the Michigan Corpus of Upper-Level Student Papers (MICUSP)," *Corpora*, 8: 183–207.

Hardy, J. A. and E. Friginal (2012), "Filipino and American Online Communication and Linguistic Variation," *World Englishes*, 31 (2): 143–61.

Helt, M. (2001), "A Multi-Dimensional Comparison of British and American Spoken English," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 171–84, Harlow: Longman.

Hymes, D. (1974), *Foundations in Sociolinguistics: An Ethnographic Approach*, Philadelphia: University of Philadelphia Press.

Hymes, D. (1984), "Sociolinguistics: Stability and Consolidation," *International Journal of the Sociology of Language*, 45: 39–45.

Irvine, J. (1979), "Formality and Informality in Communicative Events," in J. Baugh and J. Sherzer (eds), *Language in Use: Readings in Sociolinguistics*, 211–28, Englewood Cliffs, NJ: Prentice-Hall.

Jang, S.-C. (1998), "Dimensions of Spoken and Written Taiwanese: A Corpus-Based Register Study," PhD dissertation, University of Hawaii, Honolulu.

Kanoksilapatham, B. (2007), "Rhetorical Moves in Biochemistry Research Articles," in D. Biber, U. Connor and T. A. Upton (eds), *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, 73–120, Amsterdam/Philadelphia, PA: John Benjamins.

Kim, Y.-J. and D. Biber (1994), "A Corpus-Based Analysis of Register Variation in Korean," in D. Biber and E. Finegan (eds), *Sociolinguistic Perspectives on Register*, 157–81, Oxford: Oxford University Press.

Kodytek, V. (2007), *On the Replicability of the Biber Model: The Case of Czech*. Unpublished manuscript.

Kruger, H. and B. Van Rooy (2016), "Constrained Language: A Multidimensional Analysis of Translated English and a Non-Native Indigenised Variety of English," *English World-Wide*, 37 (1): 26–57.

Lamb, W. (2008), *Scottish Gaelic Speech and Writing: Register Variation in an Endangered Language*, Belfast: Cló Ollscoil na Banríona.

Longacre, R. E. (1976), *An Anatomy of Speech Notions*, Lisse: Peter de Ridder Press.

Nesi, H. and S. Gardner (2012), *Genres across the Disciplines: Student Writing in Higher Education*, Cambridge: Cambridge University Press.

Ochs, E., T. Givón and J. Sadock (1979), "Planned and Unplanned Discourse," in T. Givón (ed), *Discourse and Syntax*, 51–80, New York: Ablex.

Parodi, G. (2007), "Variation across Registers in Spanish: Exploring the El-Grial Pucv Corpus," in G. Parodi (ed), *Working with Spanish Corpora*, 11–53, London: Continuum.

Perez-Paredes, P. and M. Sanchez-Tornel (2015), "A Multidimensional Analysis of Learner Language during Story Reconstruction in Interviews," in M. Callies and S. Gôtz (eds), *Learner Corpora in Language Testing and Assessment*, 141–62, Amsterdam/Philadelphia, PA: John Benjamins.

Purvis, T. (2008), "A Linguistic and Discursive Analysis of Register Variation in Dagbani," PhD dissertation, Indiana University, Bloomington, IN.
Quaglio, P. (2009), *Television Dialogue: The Sitcom Friends vs. Natural Conversation*, Amsterdam/Philadelphia, PA: John Benjamins.
Reppen, R. (2001), "Register Variation in Student and Adult Speech and Writing," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 187–99, Harlow: Longman.
Rey, J. (2001), "Changing Gender Roles in Popular Culture: Dialogue in Star Trek Episodes from 1966 to 1993," in S. Conrad and D. Biber (eds), *Variation in English: Multi-Dimensional Studies*, 138–56, Harlow: Longman.
Sáiz, M. (1999), "A Cross-Linguistic Corpus-Based Analysis of Linguistic Variation," PhD dissertation, UMIST, Manchester.
Titak, A. and A. Roberson (2013), "Dimensions of Web Registers: An Exploratory Multi-Dimensional Comparison," *Corpora*, 8 (2): 235–60.
Ure, J. (1982), "Introduction: Approaches to the Study of Register Range," *International Journal of the Sociology of Language*, 35: 5–23.
Van Rooy, B. (2008), "A Multidimensional Analysis of Student Writing in Black South African English," *English World-Wide*, 29 (3): 268–305.
Van Rooy, B. and L. Terblanche (2009), "A Multidimensional Analysis of a Learner Corpus," in A. Renouf and A. Kehoe (eds), *Corpus Linguistics: Refinements and Reassessments*, 239–54, Amsterdam: Rodopi.
Van Rooy, B., L. Terblanche, C. Haase and J. Schmied (2010), "Register Differentiation in East African English: A Multidimensional Study," *English World-Wide*, 31: 311–49.
Veirano Pinto, M. (2014), "Dimensions of Variation in North American Movies," in T. Berber Sardinha and M. Veirano Pinto (eds), *Multi-Dimensional Analysis, 25 Years On: A Tribute to Douglas Biber*, 109–49, Amsterdam/Philadelphia, PA: John Benjamins.
Xiao, R. (2009), "Multidimensional Analysis and the Study of World Englishes," *World Englishes*, 28 (4): 421–50.
Xiao, Z. and T. McEnery (2005), "Two Approaches to Genre Analysis: Three Genres in Modern American English," *Journal of English Linguistics*, 33 (1): 62–82.
Zhang, M. (2016), "A Multidimensional Analysis of Metadiscourse Markers across Written Registers," *Discourse Studies*, 18 (2): 204–22.
Zhang, Z.-S. (2012), "A Corpus Study of Variation in Written Chinese," *Corpus Linguistics and Linguistic Theory*, 8 (1): 209–40.