# 17

# What can a corpus tell us about registers and genres?

*Bethany Gray*

## 1 Two perspectives on text varieties: register and genre

In a typical day, we encounter a range of receptive and productive language events – having conversations with our family, reading a novel, writing emails to colleagues, revising a paper for a class, reading news articles online, interacting with the server at a coffee shop…and many more. Each of these language events varies in terms of the situations in which they occur. The most obvious difference is mode (speech versus writing), but other situational differences also exist, for example:

- Purpose: e.g. news articles are informational, a novel entertains and text messages are interpersonal;
- The participants and their relationships: e.g. close personal relationships between family members, a student producing an assignment following a teacher's requirements, a knowledgeable writer of news articles writing for the general public;
- Degree of interactiveness: e.g. face-to-face service encounters and emails involve direct interaction between participants, but one occurs in real-time while the other may be asynchronous, whereas there is no direct interaction between the writer and reader of a news article.

These situational differences are associated with particular linguistic and rhetorical patterns of use. For example, novels are often written in the past tense, accompanied by many adjectives and adverbs, to narrate a story and help the reader create a mental picture of the events and characters. A service encounter at a coffee shop typically initiates with a greeting and a question from the barista asking what you'd like. In brief, we expect particular patterns of language use with each type of communicative event that we take part in.

These different types of communicative events, generally called text "categories" or "varieties", are directly relevant for corpus-based linguistics. Consider the following definition of a "corpus":

DOI: 10.4324/9780367076399-17

> a large and principled sample of texts designed to represent *a target domain of language use* (e.g. a language, dialect, or register).
>
> *(emphasis added, Egbert et al. forthcoming)*

Each of the varieties listed earlier can be considered a domain of language use. Many corpora are introduced with the phrase "a corpus of…", followed by a labeling of the domain it is intended to represent – a corpus *of business emails*, a corpus *of university lectures*, a corpus *of argumentative essays produced by L2 writers*, etc. Even general corpora are usually structured as sub-corpora of particular text categories. For example, the Corpus of Contemporary American English (COCA) is structured around sub-corpora of unscripted television/radio programs, fiction, popular magazines, newspapers and academic journals (Davies 2009), with recent additions of blogs and scripted TV/movies.

This recognition of text varieties has long been an integral component of corpus-based research; decades of empirical research has demonstrated that linguistic patterns of use are strongly mediated by text variety (Biber and Conrad 2019). Biber (2010: 243) goes so far as to claim that this research has 'shown that descriptions of grammatical variation and use are not valid for the language as a whole'. Thus, corpus-based research which fails to account for text categories in a systematic way is unlikely to produce meaningful descriptions of language use. Consider the frequency of a single part-of-speech category: nouns. Biber *et al.* (1999: 65) report that nouns occur in news writing at a rate of about 300,000 times per 1 million words. In contrast, nouns occurred in conversation just over 100,000 times per million words. Had the analysis presented a single rate of occurrence for "English" (about 200,000 times per million words if the corpora were combined), that pattern of use would fail to reflect reality. Recognising this inherent variation across text categories, corpus research has sought to uncover the typical patterns of particular domains/text categories and to map the variation that occurs across the range of domains that we encounter.

Given the integral role of text categories in corpus linguistics, it is not surprising that various terms have been used to refer to such categories, such as "register" and "genre". As Biber and Conrad (2019: 21–3) point out, some scholars use "genre" or "register" exclusively, others use the terms interchangeably and still others differentiate between the two. Of those which differentiate register and genre, readers may be most familiar with the systemic functional linguistics (SFL) approach. In SFL, "register" and "genre" exist on different semiotic planes (the "context of situation" and the "context of culture", respectively; see Halliday 1978; Hasan 1995).

However, for our purposes in relating the concept of text categories to corpus linguistics, "register" and "genre" can both be considered terms to refer to different domains of language use. Instead, the remainder of this chapter will follow Biber and Conrad (2019) in distinguishing between register and genre *approaches* to the analysis of text categories. That is, the chapter takes the position that any text category can be analysed from a register or a genre perspective.

In a register perspective, the analyst is concerned with the pervasive linguistic features that characterize the text variety and the connections that can be drawn between observed linguistic patterns of use and the non-linguistic characteristics of the variety:

> The underlying assumption of the register perspective is that core linguistic features (e.g. pronouns and verbs) serve communicative functions. As a result, some

> linguistic features are common in a register because they are functionally adapted to the communicative purposes and situational context of texts from that register.
> *(Biber and Conrad 2019: 2)*

Studies taking a register perspective often focus on lexical, phraseological, grammatical and lexico-grammatical features, documenting their quantitative distributions, lexical associations and discourse functions in specific registers or across registers. For example, the *Longman Grammar of Spoken and Written English* (Biber *et al.* 1999; republished as Biber *et al*. 2021) found that *wh*-clefts (e.g. *what I'd like to do is…, what this shows is…*) are more common in speech than in written registers, providing a way to manage the flow of information. Deroey (2012) then analysed the discourse functions of such clefts in academic lectures, finding that 67 per cent were used for an "informing" purpose (describing, interpreting or providing a gloss for concepts introduced in the lecture), while 15 per cent were used to organise the classroom discourse – two functions that are particularly important given the spoken mode and informational purpose of academic lectures. The register approach will be described in Sections 2 and 3.

In contrast, a genre perspective of text varieties is concerned with text structure and the conventionalised components that combine to create a text, which are typical or expected of a particular text variety (Biber and Conrad 2019: 2). For example, business emails conventionally begin with a greeting/salutation, and television game shows typically include a segment in which the players are introduced. Corpus-based research has less commonly focused on these sorts of conventionalised structures, more often addressing questions of text structure. One line of research in particular has demonstrated the contributions of the corpus-based genre perspective: rhetorical move analysis. Thus, genre-based approaches are the focus of Section 4.

## 2 Corpus-based register studies

Corpus-based research on the linguistic characteristics of registers has resulted in a large canon of empirical research findings. This section provides an overview of the major trends and highlights several topics that have received considerable attention (for a review of many earlier register studies, see Biber 2010). The studies discussed here share a goal of describing the quantitative distributions and functional uses of linguistic features in particular discourse contexts. However, these studies also demonstrate the breadth of corpus-based register research in terms of several parameters:

1. The scope of the registers investigated, from studies that directly compare a range of spoken and written registers, to those focusing on registers within speech or writing, to studies that offer detailed descriptions of a single register (e.g. Handford 2010 on business meetings).
2. The nature of the registers investigated, with some studies focusing on relatively general registers (e.g. conversation or academic writing) and others analysing more specialised domains (e.g. academic lectures or historical fiction novels).
3. The linguistic level of the features being analysed, including lexical, phraseological, grammatical/syntactic and lexico-grammatical features, or a constellation of features that form a functional "system" (such as stance, the expression of epistemic certainty/doubt, attitudes and evaluations).

4. The overarching goal of the research: to document the use and variability of a particular linguistic feature as mediated by register, or to describe registers with respect to particular linguistic features.

Research with a primary focus on linguistic features often accounts for register in systematic ways, recognising that linguistic patterns are variable across registers. For example, two major corpus-based research grammars have documented the variable patterns of use of grammatical features across registers in terms of quantitative distributions, their typical lexical associations and discourse functions. *The Longman Grammar of Spoken and Written English* (*LGSWE,* Biber *et al.* 1999) documents patterns of use across four general registers (conversation, fiction, news and academic writing), and the *Cambridge Grammar of English* (Carter and McCarthy 2006) accounts for spoken and written variation.

The remainder of this section is organised around the linguistic level of the analysis, focusing specifically on how these features are examined relative to register.

## *Phraseological patterns*

Register research on phraseological patterns has revealed fundamental differences in the frequency and nature of formulaic language across registers. While Greaves and Warren (Chapter 15, this volume) provide a broader overview of different types of multi-word units, here I focus specifically on two types of phraseological patterns: lexical bundles and frames (or p-frames).

The term "lexical bundle" refers to frequent, recurrent continuous sequences identified empirically based on corpus analysis. Biber *et al.* (1999) identify major patterns of register variation, showing that the most common sequences in speech and writing differ in terms of their typical structures and functions. Biber *et al.* (2004) further develop the framework, finding that:

1. Two spoken registers (academic speech and conversation) generally contain more lexical bundles than written registers (textbooks and academic prose);
2. Bundles differ structurally and functionally across registers: written registers use informational/referential noun and prepositional phrase bundles (e.g. *as a result of*), while conversation relies primarily on dependent clauses (e.g. *you might want to*) and verb-based bundles (e.g. *take a look at*) conveying stance;
3. Classroom teaching integrated the more informational, noun/prepositional phrase bundles and discourse-organising bundles alongside the bundle types typical of speech, which they relate to the instructional/informational purpose of that register.

A common goal has been to identify the most important phraseological units in specific registers. Academic writing has received particular attention, such as Candarli and Jones (2019) on research articles in education, Cortes (2004) on student and published writing in history and biology and Hyland (2008) on bundles across disciplines. Bundle use has also been explored in spoken academic language (e.g. Nesi and Basturkmen 2006) and in a range of non-academic registers (e.g. Breeze 2013 on legal registers; Mahlberg *et al.* 2019 on nineteenth-century novels).

Frames (or p-frames) are a related phraseological feature that have shown clear register-based patterns. Frames are frequently recurring sequences of words with an

internal variable "slot" that can be realised with multiple words (e.g. *the * of the* can be realised as *the end of the* and *the purpose of the*). Such sequences are seen as highly productive building blocks of discourse, and the frequency and nature of frames vary across speech and writing. Gray and Biber (2015) found a higher frequency of frames in conversation than in academic writing, with frames that were more "fixed" in nature (i.e. they had a limited number of different words occurring in the variable slots). Meanwhile, frames in academic writing were more likely to be variable (occurring with a wider range of fillers). Frames have also been investigated in specialised registers, such as book reviews (Römer 2010), pedagogical materials for engineering courses (Nekrasova-Beker 2019) and L2 English writing (Garner 2016).

### *Grammatical and lexico-grammatical features*

At the most general level, grammatical and lexico-grammatical studies of register variation have focused on the speech–writing divide, often focusing on individual linguistic features. For example, Diani (2008) analyses emphasisers in spoken and written academic registers, while Kennedy (2002) examines the distribution of modal verbs across registers in the BNC. These features can be quite narrowly defined, such as Lindley's (2020) study of progressive aspect verb phrases occurring with *always* in six spoken and written registers, or more broadly, such as Biber's (1999) study on *that*- and *to*-complement clauses in speech and writing.

Other research has focused on spoken or written registers exclusively, such as Barbieri's (2005) comparison of quotatives across four spoken registers (conversation, service encounters, study groups and office hours) and Nesselhauf and Römer (2007) on the lexical associations of progressive aspect in speech. Within written registers, comparisons across registers are common (e.g. Zhang 2015 on extraposed clauses across academic versus popular writing; Crompton 2017 on the use of demonstrative *this* in anaphoric reference in three argumentative registers). Analyses of individual registers are also prevalent, such as Padula *et al.* (2020) on popularised engineering articles.

While this research represents a very small sample of register-based grammar research, these studies illustrate the focus on detailed analyses of particular grammatical features that consider patterns of use in specific registers – including quantitative distributions, frequent lexical associations and the discourse functions of the features within particular register contexts.

It is also possible to take a slightly different perspective, in which the focus of corpus-based register research is not on a single individual feature, but rather on a set of related features, thus enabling descriptions of registers in terms of broader linguistic patterns. One such area of research considers structural devices related to grammatical complexity. This research has documented fundamentally distinct patterns across registers, with spoken registers relying on grammatical devices that are clausal in nature (e.g. verb + *that*-complement clauses), while complexity in informational written registers is largely the result of phrasal structures (e.g. attributive adjectives, nouns as noun modifiers and prepositional phrases) (Biber 1988). Complexity has been investigated synchronically and diachronically across spoken and written registers (Biber and Gray 2016), as well as in more specialised written registers (e.g. military/navy language in Noguera-Diaz and Perez-Paredes 2019) and student writing (e.g. Staples *et al.* 2016).

### *Stance and evaluation*

A second area of research that involves a set of related linguistic features is the expression of stance and evaluation – the speaker/writer's marking of epistemic certainty/doubt, attitudes and evaluations. Biber *et al.* (1999: Chapter 12) survey a range of lexicogrammatical markers of stance (e.g., modal verbs, stance adverbials, *that*- and *to*-complement clauses controlled by stance nouns, verbs and adjectives), showing distinctive patterns across spoken and written registers in terms of the overall marking of stance (with a higher preponderance of stance in speech), the typical stance meanings (a focus on epistemic meanings in academic writing) and the structural devices most frequently used for stance expression across registers.

Epistemic stance has been viewed as an important component of academic writing. Hyland's (1998) study of boosting (expressions of certainty) and hedging (expressions of doubt) across disciplines has been particularly influential, leading to detailed investigations of particular stance devices, such as stance noun + *that*-clauses (Charles 2007; Jiang 2017). Recent developments have focused on expanding definitions of stance to account for the more implicit marking of stance in written registers (e.g. Biber and Zhang 2018) and re-envisioning the issue of formality in stance marking across registers (Larsson 2019).

### *Applied register research: comparing pedagogical registers to target registers*

Many of the studies in this section have clear relevance to the teaching and learning of academic registers, providing detailed descriptions of particular registers important for novice and learner language users (see also Coxhead, this volume). Noteworthy from a register perspective is research that compares domains used for language instruction with target, "real-life" registers. For example, TED talks have been used to teach academic listening and notetaking, on the assumption that they are similar to academic lectures (spoken and informational). However, TED talks are typically shorter, more scripted, lack direct interaction with the audience and are intended for entertainment as well as information. Thus, studies have investigated the extent to which TED talks and lectures share linguistic patterns of use in terms of lexical density, speech rate and academic vocabulary use (Wingrove 2017) and lexical bundles (Liu and Chen 2019). Another example of this type of comparison is Miller's (2011) analysis of reading passages from English for Academic Purposes (EAP) textbooks and university textbooks. The goal of such research is to evaluate whether registers used to simulate a target domain for pedagogical purposes are indeed linguistically similar to that target register. This research has consistently demonstrated linguistic similarities and differences between the domains being used for pedagogical input and the target registers they are intended to help students learn, to better inform how these resources are utilised in the language classroom.

Corpus-based studies of registers produced by novice language users and L2 learners are common (see Pérez-Paredes and Mark; Römer, this volume). One line of research has compared novice or L2 language production to the target registers they will need to produce. For example, Conrad (2018) analyses the use of passive voice in student writing in civil engineering with reports written by practicing engineers, finding that students used passive voice much more frequently, whereas practitioners avoided the use of passive voice.

Other research has focused directly on how register mediates novice and learner language production. For example, Staples and Reppen (2016) analyse a range of lexico-grammatical devices in two first-year composition registers (rhetorical analysis and argumentative papers) produced by L2 English writers, while Staples *et al.* (2016) analyse grammatical complexity features in university student writing across levels, disciplines and registers (essays, critiques, case studies and explanation papers). Such research has shown that register variation is systematic even in learner and novice writing, further supporting the need to account for register in learner corpus research.

## 3 Uncovering register variation through multi-dimensional analysis

The studies surveyed in the previous section had the goal of investigating characteristics of a particular feature (or set of related features) in a limited set of registers. In this section, we turn to research with a different goal: to model overall patterns of register variation in terms of a comprehensive set of linguistic features.

Multi-dimensional (MD) analysis was developed during the 1980s (e.g. Biber 1988) to enable comprehensive, comparative descriptions of spoken versus written registers (see Biber 2019 for a discussion of the theoretical precursors of the MD approach). The MD analysis approach involves carrying out exploratory factor analysis to identify patterns of co-occurrence between a wide range of linguistic features (e.g. semantic categories of words, grammatical features and part-of-speech classes, lexcio-grammatical combinations and syntactic features). The statistical analysis results in multiple "dimensions", or sets of co-occurring features, which can be used to quantitatively describe each text and sub-corpus. Finally, each dimension is interpreted functionally.

To illustrate, consider an example from Gray (2015), an MD analysis of quantitative, qualitative and theoretical research articles in six disciplines that revealed four dimensions of variation. Figure 17.1 shows the quantitative characterisation of each sub-corpus (i.e. its dimension score) along Dimension 2; the two sets of features that statistically co-occurred in the data on this dimension are also presented. Sub-corpora on the positive end of the dimension exhibited relatively higher frequencies of the features associated with this pole of the dimension, while sub-corpora on the negative end had higher frequencies of the complementary set of features.

Based on this distribution of the registers/disciplines, the underlying communicative functions of the co-occurring features and analyses of text excerpts, Gray (2015) interpreted the positive pole of this dimension as showing narrative description. Features such as third person pronouns, past tense and perfect aspect were especially common in qualitative research, as well as in humanities and social science disciplines, as they described the participants and events in the research – they told a story (select narrative features **bolded** or underlined):

(1)
*Qualitative Applied Linguistics*
In the HIST course I **observed**, all APs **began** with a topic **introduction**, followed by a rationale for topic choice. At this stage, **presenters** usually **included** personal anecdotes that **positioned themselves** as 'licensed' to discuss their topic.
*[AJRC-AL-QL]*

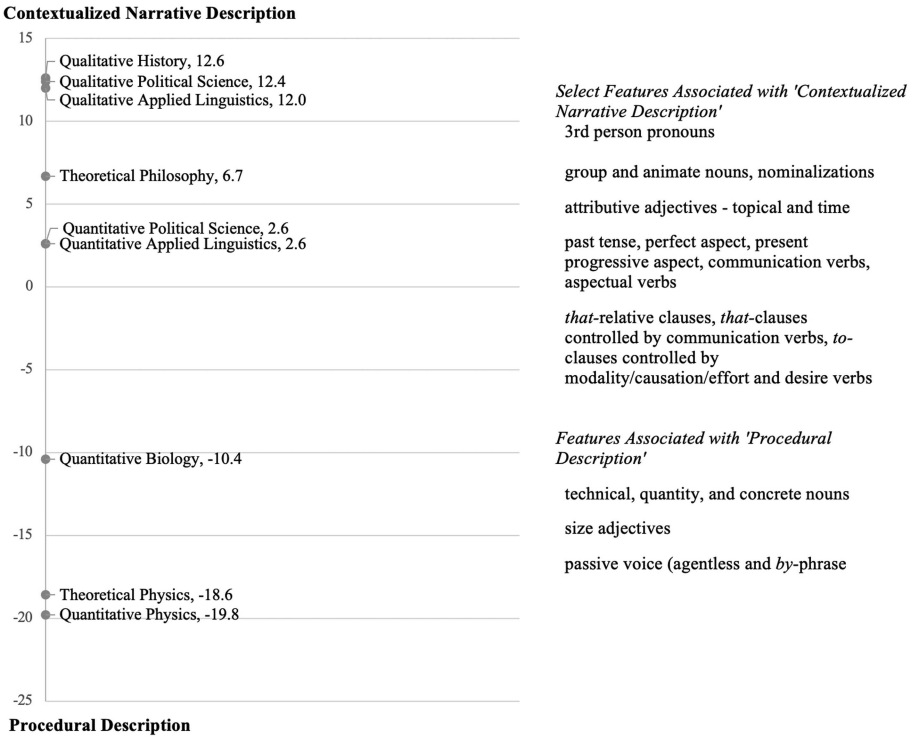


*Figure 17.1* Distribution of RA types in six disciplines along Dimension 2 "Contextualized Narrative Description" vs. "Procedural Description" (adapted from Gray 2015)

In contrast, the features associated with the negative pole (which were much more common in the hard sciences) created a procedural description of previous findings and of the actions carried out in the study procedures (select procedural features **bolded**):

(2)
*Quantitative Physics*
Many experimental features suggestive of the **liquid-gas phase** transition of nuclear **matter have been observed** in the Fermi **energy** regime of **nucleus-nucleus** collisions. […]
The **data have been obtained by** the Indra-Aladin Collaborations (Indra at GSI). The Indra set up **is described** in Refs. [21,22].

*[AJRC-PHY-QT]*

This brief example illustrates three characteristic that distinguish MD analysis from other approaches, as identified by Biber (2019: 13). First, dimensions represent a quantitative, continuous parameter of variation, and text categories fall along this continuous range according to the extent to which they are similar and different. Second, the sets of features that represent each dimension are determined empirically (based on factor analysis). Finally, 'no single parameter or dimension is adequate in itself to capture the full range of variation among registers' (Biber 2019: 13). For example, the

MD analysis described earlier revealed three additional dimensions of variations, with each dimension revealing a different distribution of disciplines/text types along the dimension.

Originally developed to investigate the differences between spoken and written registers, Biber's (1988) dimensions have become ubiquitous in the description of a wide range of registers in English and have revealed five parameters of variation:

1. Dimension 1: Involved vs. informational production;
2. Dimension 2: Narrative discourse;
3. Dimension 3: Situation-dependent vs. elaborated reference;
4. Dimension 4: Overt expression of persuasion;
5. Dimension 5: Impersonal style.

These original 1988 dimensions have been used to explore variation in a wide range of registers, including, TV dialogue (Al-Surmi 2012; Berber Sardinha and Veirano Pinto 2017, 2019); student, academic and practitioner writing in engineering (Conrad 2018); and historical change in speech-based and written registers (Biber and Finegan 2001).

New MD models have identified dimensions of variation in specialised workplace domains (Friginal 2008 on call center discourse; Staples *et al.* 2020 on health care interactions), spoken and written academic language (Biber 2006), disciplinary writing (Thomspon *et al.* 2017), university student writing (Hardy and Friginal 2016; Gardner *et al.* 2019), L2 English assessment data (LaFlair and Staples 2017) and pop songs (Bertoli-Dutra 2014). The MD analysis approach has also been applied to languages other than English, including Brazilian Portuguese (Berber Sardinha *et al.* 2014), Spanish (Biber *et al.* 2006) and Korean (Kim and Biber 1994), among others.

## 4 Corpus-based genre studies

As introduced in Section 1, a genre approach to analysing text varieties focuses on conventionalised structures used to construct texts, or what Biber and Conrad (2019) call 'genre features'. Biber and Conrad (2019) identify two types of genre features: (a) features that occur one time in a conventionalised location within a text (such as the opening and closing of a business letter) and (b) the rhetorical organisation or structuring of the text. Corpus-based analyses of the first type of genre feature are relatively rare, but corpus-based studies of the rhetorical structuring of texts have been more common. Perhaps the most productive line of corpus-based genre studies has employed Swalesian rhetorical move analysis, where *moves* are 'discoursal or rhetorical units performing coherent communicative functions in texts' (Swales 2004: 228–9), such as 'introduce the cause and/or establish credentials' in fundraising letters (Upton 2002).

While early move analyses were typically carried out on individual texts or a highly restricted sample (as the analysis required manual coding of moves), studies that analyse the move structure across texts in a corpus are now common, with increasingly large corpora (e.g. Cotos *et al.*'s 2016 analysis of 900 research articles). Corpus-based move analyses typically (a) use or adapt an existing move framework or develop a move/step structure for the genre under investigation; (b) code the corpora for moves and steps; (c) quantify the analyses to identify the frequency of particular moves/steps, characterising discoursal units that are obligatory and optional; and (d) examine the typical patterning of moves (i.e. the ordering of the moves in the genre of interest).

Written academic genres and disciplinary research writing have by far received the most attention. These studies often focus on specific sections of the research article (e.g. Tseng 2018 on theoretical framework sections) and on particular disciplines (e.g. Ye 2019 on energy engineering). However, additional research on other written academic registers is becoming more prevalent, such as Cotos (2019) on grant proposals and Samar *et al.* (2014) on conference abstracts.

Corpus-based move analyses have also expanded to spoken academic genres: e.g. Lee (2016) on EAP classroom lessons and Hu and Liu (2018) on three-minute thesis presentations. Move/step schema are also being developed and applied to a range of non-academic registers, such as Zhang and Vásquez (2014) on online reviews and Groom and Grieve (2019) on British patents.

## 5 Emerging perspectives on register and genre

In this section, I conclude the chapter by exploring two areas of research which are picking up momentum within the field.

### *Juxtapositions of the register and genre approaches*

One area of research that has become more common in recent years is the integration of the register and genre perspectives to language variation – that is, research which pairs analyses of rhetorical structuring and/or rhetorical conventions of texts with the pervasive linguistic characteristics of text varieties. For example, Cortes (2013) associates lexical bundles in research article introductions with three rhetorical moves, identifying bundles that seem to "trigger" particular moves. Omidian *et al.* (2018) also pair bundles with move analysis, while Yoon and Casal (2020) examine p-frames in conference abstracts.

Recent research has investigated broader sets of linguistic features, such as syntactic and lexical complexity measures (Tankó 2017) and more than 40 grammatical features (Kanoksilapatham 2007). However, these studies have been restricted in terms of the text categories and/or disciplines investigated. Most recently, Gray *et al.* (2020) have broadened the scope of this combination of genre and register approaches. Using a corpus of 900 research articles across 30 disciplines, they carry out multi-dimensional analysis on 14 moves across all sections of *Introduction-Method-Results-Discussion* (IMRD) research articles. They identify four dimensions of variation, enabling comprehensive description of the typical linguistic realisations of rhetorical moves regardless of discipline.

### *Expanded analyses of web registers/genres*

The past decade has seen an explosion in the amount of communicative events that we experience through electronic and web-based modalities, and corpus-based register and genre research has likewise expanded to provide a fuller understanding of linguistic patterning of these electronic registers. Blogs have received considerable attention; for example, Smith (2019) investigates personal blogs for their adherence to style guide prescriptivisms compared to news writing. Herring and Paolillo (2006) examine gender and genre variation in blogs, while Zou and Hyland (2019) analyse how writers create engagement in academic blogs. Research has also focused on a number of other

individual web registers, such as online question and answer forums (Baker and Egbert 2018) and online consumer reviews (Vásquez 2012).

Studies on web registers have also often relied on multi-dimensional analysis approaches, such as Liimata's (2019) MD of posts on Reddit, Ehret and Taboada's (2020) MD of online news comments (compared to conversation) and Clarke and Grieve (2017) on abusive language on Twitter. Efforts have also been undertaken to provide a fuller accounting of the range of online registers. Berber Sardinha (2018) uses MD analysis to compare blogs, webpages, Facebook posts, Twitter posts and emails. But perhaps one of the most revolutionary developments in corpus-based analyses of web registers are attempts to inductively identify the full range of public internet registers. Biber *et al.* (2015) create large random samples of texts from the searchable Web and develop a taxonomy of text categories to capture the traditional and new registers that can be found online. Biber and Egbert (2018) then provide a comprehensive linguistic analysis of four major register categories (narrative web registers, opinion/advice/persuasion registers, informational descriptions/explanations/procedures and oral registers). One of the major findings that has emerged out this research is the degree of hybridisation that occurs in online registers, where individual texts reflect characteristics of multiple registers.

## Further reading

Biber, D. and Conrad, S. (2019) *Register, Genre, and Style*, 2nd edn, Cambridge: Cambridge University Press. (Part 1 of this book explores three perspectives on text varieties and introduces a practical framework for corpus-based register studies. Part II synthesises research on spoken, written, academic/professional and electronic registers. Part III introduces multi-dimensional analysis and contextualises register studies within the field of linguistics.)

Biber, D. and Reppen, R. (eds) (2015) *The Cambridge Handbook of English Corpus Linguistics*, Cambridge: Cambridge University Press. (Part III, titled "Corpus Analysis of Varieties", synthesises research on particular registers. Chapters include spoken discourse, academic writing, register variation, diachronic register description, literary registers, World Englishes, English as a lingua franca and learner language.)

Berber Sardinha, T. and Veirano Pinto, M. (eds) (2019) *Multi-Dimensional Analysis: Research Methods and Current Issumes*, London: Bloomsbury. (This book provides an up-to-date overview of the theory, methods and application of multi-dimensional analysis. It covers all stages of the research process, from the history and theory of the approach to corpus design and annotation (Part 1) and the statistical analyses and functional interpretation (Part 2). Part 3 contains three illustrations of the MD method.)

Römer, U., Cortes, V. and Friginal, E. (eds) (2020) *Advances in Corpus-Based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*, Amsterdam: John Benjamins. (This edited volume assembles 14 corpus-based studies of a range of written academic registers (EFL student writing, first-year composition, research articles across disciplines, stand-alone-literature reviews, conference abstracts). Studies also demonstrate a range of linguistic focci: academic vocabulary, lexical bundles and p-frames, verb constructions, adjectives as nominal pre-modifiers and multi-dimensional analyses.)

Staples, S. (2015) *The Discourse of Nurse-Patient Interactions*, Amsterdam: John Benjamins. (This book presents a comprehensive analysis of a professional spoken register and interactions between US (L1 English) and international (L2 English) nurses and patients in a US context. The analyses encompass situational analyses of the register, accompanied by lexico-grammatical descriptions and the analysis of spoken features (fluency, prosody and non-verbal communication).

## References


Al-Surmi, M. (2012) 'Authenticity and TV Shows: A Multidimensional Analysis Perspective', *TESOL Quarterly* 46(3): 472–95.

Baker, P. and Egbert, J. (eds) (2018) *Triangulating Methodological Approaches in Corpus Linguistic Research*, London: Routledge.

Barbieri, F. (2005) 'Quotative Use in American English: A Corpus-Based, Cross-Register Comparison', *Journal of English Linguistics* 33(3): 222–56.

Berber Sardinha, T. (2018) 'Dimensions of Variation across Internet Registers', *International Journal of Corpus Linguistics* 23(2): 125–57.

Berber Sardinha, T., Kauffmann, C. and Acunzo, C. M. (2014) 'A Multi-Dimensional Analysis of Register Variation in Brazilian Portuguese', *Corpora* 9(2): 239–71.

Berber Sardinha, T. and Veirano Pinto, M. (2017) 'American Television and Off-Screen Registers: A Corpus-Based Comparison', *Corpora* 12(1): 85–114.

Berber Sardinha, T. and Veirano Pinto, M. (2019) 'Dimensions of Variation across American Television Registers', *International Journal of Corpus Linguistics* 24(1): 3–32.

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*, Amsterdam: John Benjamins, pp. 149–76.

Biber, D. (1988) *Variation across Speech and Writing*, *Cambridge*: *Cambridge University Press*.

Biber, D. (1999) 'A Register Perspective on Grammar and Discourse: Variability in the Form and Use of English Complement Clauses', *Discourse Studies* 1(2): 131–50.

Biber, D. (2006) *University Language: A Corpus-Based Study of Spoken and Written Language*, Amsterdam: John Benjamins.

Biber, D. (2010) 'What Can a Corpus Tell us about Registers and Genres?', in A. O'Keeffe and M. J. McCarthy (eds)*The Routledge Handbook of Corpus Linguistics*, 1st edn, London: Routledge, pp. 241–54.

Biber, D. (2019) 'Multi-Dimensional Analysis: A Historical Synopsis', in T. Berber-Sardinha and M. Veirano Pinto (eds) *Multidimensional Analysis: Research Methods and Current Issues*, London: Bloomsbury, pp. 11–26.

Biber, D. and Conrad, S. (2019) *Genre, Register, and Style*, 2nd edn, Cambridge: Cambridge University Press.

Biber, D., Conrad, S. and Cortes, V. (2004) '"*If You Look At*…": Lexical Bundles in University Teaching and Textbooks', *Applied Linguistics* 23: 371–405.

Biber, D., Davies, M., Jones, J. and Tracy-Ventura, N. (2006) 'Spoken and Written Register Variation in Spanish: A Multi-Dimensional Analysis', *Corpora* 1(1): 1–37.

Biber, D. and Egbert, J. (2018) *Register Variation Online*, Cambridge: Cambridge University Press.

Biber, D., Egbert, J. and Davies, M. (2015) 'Exploring the Composition of the Searchable Web: A Corpus-Based Taxonomy of Web Registers', *Corpora* 10(1): 11–45.

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

Biber, D. and Gray, B. (2016) *Grammatical Complexity in Academic English: Linguistic Change in Writing*, Cambridge: Cambridge University Press.

Biber, D. and Zhang, M. (2018) 'Expressing Evaluation without Grammatical Stance: Informational Persuasion on the Web', *Corpora* 13(1): 97–112.

Biber, D., Johansson, S., Leech, G, Conrad, S. and Finegan, E. (1999) *The Longman Grammar of Spoken and Written English*, London: Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (2021) *The Grammar of Spoken and Written English*, Amsterdam: John Benjamins.

Breeze, R. (2013) 'Lexical Bundles across Four Legal Genres', *International Journal of Corpus Linguistics* 18(2): 229–53.

Candarli, D. and Jones, S. (2019) 'Paradignmatic Influences on Lexical Bundles in Research Articles in the Discipline of Education', *Corpora* 14(2): 237–63.

Carter, R. A. and McCarthy, M. J. (2006) *Cambridge Grammar of English*, Cambridge: Cambridge University Press.

Charles, M. (2007) 'Argument or Evidence? Disciplinary Variation in the Use of the Noun *that* Pattern in Stance Construction', *English for Specific Purposes* 26: 203–18.
Clarke, I. and Grieve, J. (2017) 'Dimensions of Abusive Language on Twitter', *Proceedings of the* First Workshop on Abusive Language Online, 1–10.
Conrad, S. (2018) 'The Use of Passives and Impersonal Style in Civil Engineering Writing', *Journal of Business and Technical Communication* 32(1): 38–76.
Cortes, V. (2004) 'Lexical Bundles in Published and Student Disciplinary Writing: Examples from history and biology', *English for Specific Purposes* 23: 397–423.
Cortes, V. (2013) 'The Purpose of This Study Is to: Connecting Lexical Bundles and Moves in Research Article Introductions', *Journal of English for Academic Purposes* 12: 33–43.
Cotos, E. (2019) 'Move Analysis of Broader Impact Statements in Grant Applications', *English for Specific Purposes* 54: 15–34.
Cotos, E., Link, S. and Huffman, S. (2016) 'Studying Disciplinary Corpora to Teach the Craft of Discussion', *Writing and Pedagogy* 8(1): 33–64.
Crompton, P. (2017) 'Complex Anaphora with *This:* Variation between Three Written Argumentative Genres', *Corpora* 12(1): 115–48.
Davies, M. (2009) 'The 385+ Million Word *Corpus of Contemporary American English (1990-2008+)*', *International Journal of Corpus Linguistics* 14(2): 159–90.
Deroey, K. (2012) '*What They Highlight Is…:* The Discourse Functions of Basic *wh*-Clefts in Lectures', *Journal of English for Academic Purposes* 11: 112–24.
Diani, G. (2008) 'Emphasizers in Spoken and Written Academic Discourse: The Case of Really', *International Journal of Corpus Linguistics* 13(3): 296–321.
Egbert, J., Biber, D. and Gray, B. (forthcoming) *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness*, Cambridge: Cambridge University Press.
Ehret, K. and Taboada, M. (2020) 'Are Online News Comments Like Face-to-face Conversation? A Multi-Dimensional Analysis of an Emerging Register', *Register Studies* 2(1): 1–36.
Friginal, E. (2008) *The Language of Outsourced Call Centers: A Corpus-Based Study of Cross-Cultural Interaction*, Amsterdam: John Benjamins.
Gardner, S., Nesi, H. and Biber, D. (2019) 'Discipline, Level, Genre: Integrating Situational Perspectives in a New MD Analysis of University Student Writing', *Applied Linguistics* 40(4): 646–74.
Garner, J. (2016) 'A Phrase-Frame Approach to Investigating Phraseology in Learner Writing Across Proficiency Levels', *International Journal of Corpus Linguistics* 2(1): 31–68.
Gray, B. (2015) *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story*, Amsterdam: John Benjamins.
Gray, B. and Biber, D. (2015) 'Lexical Frames in Academic Prose and Conversation', in S. Hoffmann, B. Fischer-Starcke and A. Sand (eds) *Current Issues in Phraseology*, Amsterdam: John Benjamins, pp. 109–34.
Gray, B., Cotos, E. and Smith, J. (2020) 'Combining Rhetorical Move Analysis with Multi-Dimensional Analysis: Research Writing across Disciplines', in U. Römer, V. Cortes and E. Friginal (eds) *Advances in Corpus-Based Research on Academic Writing*, Amsterdam: John Benjamins, pp. 138–68.
Groom, N. and Grieve, J. (2019) 'The Evolution of a Legal Genre: Rhetorical Moves in British Patent Specifications, 1711 to 1860', in T. Fanego and P. Rodríguez-Puente (eds) *Corpus-Based Research on Variation in English Legal Discourse*, Amsterdam: John Benjamins, pp. 201–34.
Halliday, M. A. K. (1978) *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*, London: Edward Arnold.
Handford, M. (2010) *The Language of Business Meetings*, Cambridge: Cambridge University Press.
Hardy, J. and Friginal, E. (2016) 'Genre Variation in Student Writing: A Multi-Dimensional Analysis', *Journal of English for Academic Purposes* 22: 119–31.
Hasan, R. (1995) 'The Conception of Context in Text', in P. Fries and M. J. Gregory (eds) *Discourse in Society: Systemic Functional Perspectives. Meaning and Choice in Language: Studies for Michael Halliday*, Norwood, NJ: Ablex, pp. 183–283.
Herring, S. and Paolillo, J. (2006) 'Gender and Genre Variation in Weblogs', *Journal of Sociolinguistics* 10(4): 439–59.

Hu, G. and Liu, Y. (2018) 'Three Minute Thesis Presentations as an Academic Genre: A Cross-Disciplinary Study of Genre Moves', *Journal of English for Academic Purposes* 35: 16–30.
Hyland, K. (2008) 'As Can be Seen: Lexical Bundles and Disciplinary Variation', *English for Specific Purposes* 27: 4–21.
Hyland, K. (1998) 'Boosting, Hedging and the Negotiation of Academic Knowledge', *Text* 18(3): 349–83.
Jiang, F. (2017) 'Stance and Voice in Academic Writing: The "noun + that" Construction and Disciplinary Variation', *International Journal of Corpus Linguistics* 22(1): 85–106.
Kanoksilapatham, B. (2007) 'Rhetorical Moves in Biochemistry Research Articles', in D. Biber, U. Connor and T. Upton (eds) *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*, Amsterdam: John Benjamins, pp. 73–120.
Kennedy, G. (2002) 'Variation in the Distribution of Modal Verbs in the British National Corpus', in R. Reppen, S. Fitzmaurice and D. Biber (eds) *Using Corpora to Explore Linguistic Variation*, Amsterdam: John Benjamins, pp. 73–90.
Kim, Y. J. and Biber, D. (1994) 'A Corpus-Based Analysis of Register Variation in Korean', in D. Biber and E. Finegan (eds) S*ociolinguistic Perspectives on Register*, Oxford: Oxford University Press, pp. 157–81.
LaFlair, G. and Staples, S. (2017) 'Using Corpus Linguistics to Examine the Extrapolation Inference in the Validity Argument for a High-Stakes Speaking Assessment', *Language Testing* 34(4): 451–75.
Larsson, T. (2019) 'Grammatical Stance Marking across Registers: Revisiting the Formal-Informal Dichotomy', *Register Studies* 1(2): 243–68.
Lee, J. (2016) '"There's Intentionality Behind It…": A Genre Analysis of EAP Classroom Lessons', *Journal of English for Academic Purposes* 23: 99–112.
Liimatta, A. (2019) 'Exploring Register Variation on Reddit: A Multi-Dimensional Study of Language Use on a Social Media Website', *Register Studies* 1(2): 269–95.
Lindley, J. (2020) 'Discourse Functions of *Always* Progressives: Beyond Complaining', *Corpus Linguistics and Linguistic Theory* 16(2): 333–61.
Liu, C.-Y. and Chen, H.-J. (2019) 'Functional Variation of Lexical Bundles in Academic Lectures and TED Talks', *Register Studies* 2(2): 176–208.
Mahlberg, M., Wiegand, V., Stockwell, P. and Hennessey, A. (2019) 'Speech-Bundles in the 19th-Century English Novel', *Language and Literature* 28(4): 326–53.
Miller, D. (2011) 'ESL Reading Textbooks vs. University Textbooks: Are we Giving our Students the Input they May Need?', *Journal of English for Academic Purposes* 10: 32–46.
Nekrasova-Beker, T. (2019) 'Discipline-Specific Use of Language Patterns in Engineering: A Comparison of Published Pedagogical Materials', *Journal of English for Academic Purposes* 41: 1–12.
Nesi, H. and Basturkmen, H. (2006) 'Lexical Bundles and Discourse Signaling in Academic Lecturers', *International Journal of Corpus Linguistics*, 11(3): 283–304.
Nesselhauf, N. and Römer, U. (2007) 'Lexical-Grammatical Patterns in Spoken English: The Case of the Progressive with Future Time Reference', *International Journal of Corpus Linguistics* 12(3): 297–333.
Noguera-Diaz, Y. and Perez-Paredes, P. (2019) 'Register Analysis and ESP Pedagogy: Noun-Phrase Modification in a Corpus of English for Military Navy Submariners', *English for Specific Purposes* 53: 118–30.
Omidian, T., Shahriari, H. and Siyanova-Chanturia, A. (2018) 'A Cross-Disciplinary Investigation of Multi-Word Expressions in the Moves of Research Article Abstracts', *Journal of English for Academic Purposes* 36: 1–14.
Padula, M., Panza, C. and Munoz, V. (2020) 'The Pronoun *This* as a Cohesive Encapsulator in Engineering Semi-Popularization Articles Written in English', *Journal of English for Academic Purposes* 44: 100828.
Quaglio, P. (2009) *Television Dialogue: The Sitcom of Friends vs. Natural Conversation*, Amsterdam: John Benjamins.
Römer, U. (2010) 'Establishing the Phraseological Profile of a Text Type: The Construction of Meaning in Academic Book Reviews', *English Text Construction* 3(1): 95–119.

Samar, R. G., Talebzadeh, H., Kiany, G. R. and Akbari, R. (2014) 'Moves and Steps to Sell a Paper: A Cross-Cultural Genre Analysis of Applied Linguistics Conference Abstracts', *Text and Talk* 34(6): 759–85.

Smith, J. (2019) 'A Comparison of Prescriptive Usage Problems in Formal and Informal Written English', unpublished PhD thesis, Ames, IA: Iowa State University.

Staples, S., Egbert, J., Biber, D. and Gray, B. (2016) 'Academic Writing Development at the University Level: Phrasal and Clausal Complexity across Level of Study, Discipline, and Genre', *Written Communication* 33(2): 149–83.

Staples, S. and Reppen, R. (2016) 'Understanding First-year L2 Writing: A Lexico-Grammatical Analysis across L1s, Genres, and Language Ratings', *Journal of Second Language Writing* 32: 17–35.

Staples, S., Venetis, M. K. and Robinson, R. (2020) 'Understanding the Multi-Dimensional Nature of Informational Language in Health Care Interactions', *Register Studies* 2(2): 241–74.

Swales, J. (2004) *Research Genres*, Cambridge: Cambridge University Press.

Tankó, G. (2017) 'Literary Research Article Abstracts: An Analysis of Rhetorical Moves and their Linguistic Realizations', *Journal of English for Academic Purposes* 27: 42–55.

Thomspon, P., Hunston, S., Murakami, A. and Vajn, D. (2017) 'Multi-Dimensional Analysis, Text Constellations, and Interdisciplinary Discourse', *International Journal of Corpus Linguistics* 22(2): 153–86.

Tseng, M.-Y. (2018) 'Creating a Theoretical Framework: On the Move Structure of Theoretical Framework Sections in Research Articles Related to Language and Linguistics', *Journal of English for Academic Purposes* 33: 82–99.

Upton, T. (2002) 'Understanding Direct Mail Letters as a Genre', *International Journal of Corpus Linguistics* 7(1): 65–85.

Vásquez, C. (2012) 'Narrative and Involvement in Online Consumer Reviews: The Case of *TripAdvisor*', *Narrative Inquiry* 22(1): 105–21.

Wingrove, P. (2017) 'How Suitable Are TED Talks for Academic Listening?', *Journal of English for Academic Purposes* 30: 79–95.

Ye, Y. (2019) 'Macrostructures and Rhetorical Moves in Energy Engineering Research Articles Written by Chinese Expert Writers', *Journal of English for Academic Purposes* 38: 48–61.

Yoon, J. and Casal, J. E. (2020) 'P-Frames and Rhetorical Moves in Applied Linguistics Conference Abstracts', in U. Römer, V. Cortes and E. Friginal (eds) *Advances in Corpus-Based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*, Amsterdam: John Benjamins, pp. 282–305.

Zhang, G. (2015) '*It is Suggested That….* or *It is Better to…?* Forms and Meanings of Subject *it*-Extraposition in Academic and Popular Writing', *Journal of English for Academic Purposes* 20: 1–13.

Zhang, Y. and Vásquez, C. (2014) 'Hotels' Responses to Online Reviews: Managing Consumer Dissatisfaction', *Discourse, Context and Media* 6: 54–64.

Zou, H. and Hyland, K. (2019) 'Reworking Research: Interactions in Academic Articles and Blogs', *Discourse Studies* 21(6): 713–33.