CHAPTER 1

Introduction

1.1 Academic research writing: One register or many?

Applied linguists have long been fascinated with the written language of academia. This interest has developed and expanded over the past few decades, in part due to the premise that much can be learned about disciplinary practices and cultures by examining academic writing: the primary means of the transmission of knowledge in academic fields. The ability to use a single, overarching term like 'academic writing' belies the complexity and range of text types that fall within this label. Academic writing is not a monolithic construct, despite the ease with which we refer to 'academic writing' in general terms. Instead, academic writing is widely regarded as a register exhibiting inherent variation. We easily recognize different types of academic writing, commonly making distinctions between writing produced by students versus professional academics, between L1 and L2 writers, and between texts produced for different purposes. For example, we recognize a range of sub-registers within the domain of 'academic writing', such as textbooks, lab reports, research monographs, conference abstracts, argumentative essays, book reviews, and research articles, to name a few. A great deal of research has been devoted to describing the language of these different texts – from vocabulary use to phraseological patterns, and from grammatical characteristics to discourse structure – with the understanding that these language features are used in distinct ways in different types of academic texts.

Likewise, there is wide recognition that written academic language varies according to discipline – that disciplines utilize linguistic resources in varied ways to construct meaning and build knowledge within their disciplinary communities. The assumption is that the language used by these disciplinary communities is distinct, just as disciplines differ in their epistemological beliefs, research practices, and knowledge structures. Such variation across disciplines is perhaps never more evident than when reading texts from an unfamiliar discipline, as even readers well-versed in research writing in their own fields may be challenged to parse exact meanings out of the words, phrases, and clauses on the page. Consider, for

example, the following excerpts. Both excerpts come from research articles published in academic journals, and even come from the same part of the text – the abstract. Is understanding the content difficult? Can you guess the disciplines to which these excerpts belong?

Excerpt 1

> This paper argues that expressivism faces serious difficulties giving an adequate account of univocal moral disagreements. Expressivist accounts of moral discourse understand moral judgments in terms of various noncognitive mental states, and they interpret moral disagreements as clashes between competing (and incompatible) attitudes. I argue that, for various reasons, expressivists must specify just what mental states are involved in moral judgment. If they do not, we lack a way of distinguishing moral judgments from other sorts of assessment and thus for identifying narrowly moral disagreements. If they do, we can construct cases of intuitively real dispute that do not rest on the theory's preferred mental states. This strategy is possible because our intuitions about moral concept-ascription do not track speakers' noncognitive states. I discuss various ways of developing this basic argument, then apply it to the work of the two most sophisticated proponents of expressivism, Allan Gibbard and Simon Blackburn. I argue that neither is successful in meeting the challenge. [Philosophy abstract; Merli 2008]

Excerpt 2

> During early embryogenesis in Caenorhabditis elegans, the ATL-1CHK-1 (ataxia telangiectasia mutated and Rad relatedChk) checkpoint controls the timing of cell division in the future germ line, or P lineage, of the animal. Activation of the CHK-1 pathway by its canonical stimulus DNA damage is actively suppressed in early embryos so that P lineage cell divisions may occur on schedule. We recently found that the rad-2 mutation alleviates this checkpoint silent DNA damage response and, by doing so, causes damage-dependent delays in early embryonic cell cycle progression and subsequent lethality. In this study, we report that mutations in the smk-1 gene cause the rad-2 phenotype. SMK-1 is a regulatory subunit of the PPH-4.1 (protein phosphatase 4) protein phosphatase, and we show that SMK-1 recruits PPH-4.1 to replicating chromatin, where it silences the CHK-1 response to DNA damage. These results identify the SMK-1PPH-4.1 complex as a critical regulator of the CHK-1 pathway in a developmentally relevant context. [Biology abstract; Kim et al. 2007]

Now, think about what features of the excerpts lead you to make that determination. Perhaps the first feature that comes to mind is content or vocabulary. Both excerpts make extensive use of lexical items that are specific to the areas of inquiry in the respective disciplines: *expressivism, moral disagreements, noncognitive mental states*, and *moral concept-asscription* for the philosophy excerpt, and *embryongenesis, DNA, embryonic cell cycle progression*, and *protein phosphatase* for biology.

Even without knowing the exact meanings of all of these items, you were likely able to make a guess about the disciplines of these two abstracts.

However, the linguistic differences between the texts go beyond the content expressed by lexical items. Perhaps you also noticed the use of personal pronouns in the texts. The philosophy abstract uses many more personal pronouns generally, and uses both third person and first person pronouns to refer to the author and some unspecified individuals – the "expressivists". The biology excerpt uses a few first person pronouns, but the reference is restricted to the authors of the texts. Or, perhaps you recognized the use of 'argue' to describe the main rhetorical function in the philosophy excerpt, contrasted with the focus on reported empirical research results in the biology text (e.g., *We recently found that…, We report that…, We show that…*). Even impressionistically, we can recognize differences in the linguistic resources that are utilized in the two texts. We recognize the simple fact that language use varies between disciplines, and this acknowledgment has provided the motivation for a wealth of research into disciplinary communication practices.

According to Bazerman (1994: 104), the underlying belief of investigations into language use in the disciplines is that "the primary product of most disciplines, and a secondary product of all, are published texts, which are taken to constitute the knowledge of the disciplines". Written discourse is so integrally connected to disciplinary knowledge that Turner (2006) actually uses the presence of disciplinary discourse in his definition of a discipline: a discipline is socially constructed, has regulatory practices, and its members practice a "rhetoric of competence".

The centrality of writing to academic culture, practice, and knowledge building has led to a great deal of research in several fields, including rhetoric and composition, applied linguistics, and English for Academic Purposes (EAP). Often, studies investigating academic writing are motivated by the desire to inform the teaching of writing to native and non-native English-speaking students, through both descriptions of professional academic writing as well as through comparisons of novice writer (native and non-native English-speaking) and expert production. However, while learning about academic writing to better inform teaching content and practices is an important aim, Bazerman (1994) points out that understanding language use in the disciplines also helps us to use language more effectively, can guide editors as they work with contributor texts, and helps provide non-specialist readers with access to the discourse of the disciplines. Describing and understanding patterns of language use in academic prose allows us to understand the disciplinary cultures and practices that they embody for a variety of purposes. To do so, academic research articles have been the focus much of the research with this goal (although disciplinary differences in student writing is gaining in focus, particularly with the development of the Michigan Corpus of Upper-Level Student Papers and the British Academic Written

English corpus; for examples, see Hardy & Römer 2013; Römer & Swales 2010; Nesi & Gardner 2012; Gardner & Nesi 2012).

Disciplinary variation in academic writing is widely recognized, but we know less about the actual patterns of linguistic variation across disciplines than we know about variation across more broadly defined registers of academic writing. Becher (1981: 110) claims that "It is fairly obvious that disciplines differ from one another, but not so obvious what the differences comprise". Although Becher was not discussing the language patterns of different disciplines specifically, the research synthesis in Sections 1.3 and 1.4 below illustrate that the sentiment also applies to documenting the linguistic characteristics of texts across disciplines. Thus, one goal of this book is to elucidate systematic patterns of variation in the linguistic structure of research articles across disciplines.

However, the aims of the present research go well beyond establishing disciplinary differences in research writing. While the focus remains on research articles, part of my goal is to problematize the classification of 'research articles' as a wholly adequate register distinction for published research writing. This goal is partially motivated by the desire to acknowledge variation *within* academic disciplines, in addition to variation that corresponds to discipline-specific norms. Disciplines clearly differ in their basic characteristics, including data sources, areas of inquiry, research methods, and epistemological belief systems. However, variation with respect to these parameters also exists *within* disciplines. For example, many disciplines in the social sciences rely on both quantitative and qualitative research paradigms to conduct academic inquiry. Little research has systematically considered how differences in research paradigm (even within a discipline) might relate to distinct patterns in the linguistic structure of texts. Cao and Hu (2014: 17) claim that "the epistemological assumptions associated with quantitative and qualitative paradigms are believed to….shape the discourse and rhetorical conventions in which empirical research is presented". Their analysis of quantitative and qualitative research articles in three disciplines showed patterns of variation both across research paradigms and across disciplines – at least for the specific feature of interest, interactive metadiscourse markers. However, it is also likely that factors beyond discipline (such as research paradigm) contribute to variation in the use of many types of linguistic features in published research articles.

For illustration, consider the following two excerpts; both are from the results section of research articles in applied linguistics (both articles were published in the same journal, to avoid any differences due to the source journal). However, excerpt 3 reports on a *qualitative* research study, while excerpt 4 reports on a *quantitative* research study. Both are empirical studies, but rely on very different research methodologies and research belief systems.

Excerpt 3

> Ultimately, both the stimulated recall sessions and the talk-in-interaction sessions of both Groups 1 and 2 indicated that all students used the L to determine the meaning of the targeted forms…Groups 1 and 2 revealed four important differences (see Table 1). First, there was a difference in the fluidity of their interactions. Pairs in Group 1 engaged in smooth, continuous interaction. They talked while reading and reviewing the passage, and while discussing the target structures. By contrast, the interactions of pairs in Group 2 were characterized by frequent pauses and fragmented interaction. The students in Group 2 often laughed nervously and looked out the window during pauses, some of which went on for nearly 2 minutes. Although the majority of the students in Group 1 verbalized their thoughts in the L, the students in Group 2 who were trying to use only the L had to translate their L thoughts into the L in order to be able to share them with their conversation partners. [Qualitative applied linguistics, reporting results; Scott & de la Fuente 2010]

Excerpt 4

> The criterion for identification of linguistic variables was based on the components of the functional trisection: function, content, and accuracy (Higgs & Clifford, 1982). Table 2 and Figure 1 show results for the comparison of means t-tests for groups of nullgainers and gainers for three linguistic variables-grammar, accuracy, and vocabulary-in addition to two metalinguistic variables-self-corrected errors and sentence repairs. The score for the norm-referenced ACTR Qualifying Grammar Test, administered prior to the students' departure for Russia, is a classic variable based on domain-specific knowledge. The score reveals the percentage of correctly answered questions on the test. The scores for the group of nullgainers ranged from 36 to 78, and for the group of gainers, scores ranged from 47 to 80. Means for the nullgainers and gainers were 53.3 and 64, respectively. The result for the test of equality of group means was statistically significant at p =.042. [Quantitative applied linguistics, reporting results; Golonka 2006]

Despite coming from the same discipline, the same source journal, and even the same section of the research article, there are marked differences in the way that the outcomes of the research are being reported. The use of third person pronouns, past tense, and relative clauses create a rich, narrative-like description in excerpt 3. In contrast, excerpt 4 uses a mix of past and present tense, relatively few verbs, and many complex noun phrases – which all create a sense of succinct, procedural discourse. The potential for linguistic variation associated with factors such as research paradigm is quite likely, but as of yet these variables are largely unexplored as productive explanations of *within* discipline variation.

At the same time, very different disciplines may use similar research methodologies to explore phenomena in their respective fields. For example, both physics and philosophy rely on theoretical research to develop hypotheses and theories,

and build knowledge within the disciplines. While there are likely substantial linguistic differences due to the nature of the two disciplines, might there also be similarities due to the theoretical nature of the arguments being made? Thus, recognition of multiple types of research reports within disciplines, and how they relate to the range of research types in other disciplines, provides an opportunity to better explain the similarities and differences that can be observed *across* disciplines.

Thus, although substantial research has focused on disciplinary differences across research articles, little attention has been paid to the possibility that research articles themselves are not a monolithic concept. That is, substantial linguistic variation may exist even within the register of *research article* due to situational factors not previously considered, including the research design/methodology being reported, the nature of data, and the role of discipline. Thus, the primary goal of this book is to investigate the linguistic characteristics of registers published in academic journals both *across* and *within* disciplines, while taking into account the varied realizations of research articles (what I will call 'registers' or 'academic journal registers') in fundamentally diverse disciplines.

1.1.1 A note on 'register'

Terms like 'register,' 'genre,' 'text type,' and 'style' have long been used with a variety of meanings to categorize texts (for a comprehensive review on the uses of these terms, see Lee 2001; see also Biber 2006: 10–12). That is, linguists have sought for a way to group texts of a similar nature, and these terms have been used to describe categories of texts. The ability to create such groupings of texts has become particularly important as the field of corpus linguistics has grown. With increasing concern and awareness of the linguistic differences that exist between varieties of language has come the desire to systematically study that linguistic variation. This, in turn, results in the need for the ability to design both balanced and representative corpora that accurately characterize the language varieties under investigation.

While the term 'register' has been used widely by research coming from a variety of different theoretical frameworks, I use 'register' in the sense defined by Biber & Conrad (2009: see Chapter 1; also see Biber 1994, Biber & Finegan 1994: Chapter 1). That is, a register is a variety of language that can be characterized by the situations in which it is used. We can consider such situational factors like mode (written versus spoken), purpose (e.g., argumentative versus informative versus aesthetic), participants and the relationship between participants in a communicative event, and so on. Such factors are characteristics of the texts that are separate from the linguistic structure of those texts; however, the premise behind a

register perspective on language use is that these non-linguistic characteristics are correlated with, or associated with, the linguistic structure of those texts.

Registers can be defined at varying levels of specificity. For example, 'academic writing' is a commonly investigated register. However, within that broad register of academic writing, both textbooks and research articles can be identified as more narrowly defined registers (sometimes called 'sub-registers') under the umbrella term *academic writing*. The two registers share the overarching informational purpose of academic writing, but other non-linguistic features distinguish them from each other. For example, textbooks are intended to introduce the non-expert reader to a field, topic, or discipline, and typically cover a wide range of topics, and summarize established knowledge in the field. Research articles, on the other hand, present more focused, specialized information about new, developing knowledge in the field to expert readers.

Likewise, there is variation in the situational characteristics of texts *within* the register of 'research articles'. Research articles differ in terms of the discipline within which they fall, the journals they occur in, who their authors are, the types of methods that are used to conduct the research being reported on, and so on (see Chapters 2 and 3 for elaboration). Because of the ability to describe differences in the situational characteristics of different types of research articles, I use the terms 'register', 'sub-register', or 'academic journal register' to refer to the range of research article types. For ease of use, the majority of the time I will simply use the term 'register'. My motivation for taking this fine-grained approach to research articles is that in discussions to date (see discussion in Sections 1.2 and 1.3), research articles are generally grouped into a single register that is defined as texts which report on research and are published in academic journals. This rather coarse definition means that similarities across articles which report on distinct research methodologies and disciplines are assumed but not often investigated. In reality, the differing nature of research may be leading to substantial linguistic variation not being captured by the studies that have been conducted to date, as illustrated by the text excerpts presented above. In this book, I identify and examine sub-registers within the broader register of published academic research articles.

1.1.2 Goal of the present book

Thus, the goal of this book is to investigate the linguistic characteristics of registers published in academic journals, taking into account the varied realizations of research articles in fundamentally diverse disciplines. That is, the study seeks to go beyond the traditional and often one-dimensional analysis of a generically-defined research article to first distinguish between different types of articles (or registers)

within and across disciplines, and then to describe those registers according to their characteristic linguistic and non-linguistic features.

The research in this book is based on the Academic Journal Register Corpus, a corpus of 270 research articles from six disciplines: philosophy, history, political science, applied linguistics, biology, and physics. Research articles within these disciplines are further categorized by journal register: theoretical, qualitative, and quantitative research articles. This volume presents analyses of both the non-linguistic and linguistic characteristics of the corpus. In the analysis of the non-linguistic features of these texts, a framework for describing the situational characteristics is first proposed and then applied to each text in the corpus. The linguistic analysis relies on quantitative and qualitative analyses of data extracted through several specialized computer programs, and includes a grammatical survey of the distributions of core grammatical features, a description of grammatical features which 'elaborate' and 'compress' discourse, and a statistical analysis that identifies co-occurrence patterns of 70 linguistic features.

1.2 The linguistic characteristics of academic writing

There is a general consensus, even outside the academic community, that academic writing has distinct characteristics that set it apart from other types of language. Much research (e.g., Biber 1988; Biber et al. 1999; Biber & Gray 2010, 2016; Halliday 2004; Banks 2005, 2008; Fang, Schleppegrell & Cox 2006) has focused on describing one of the defining characteristics of academic prose: its dense reliance on nouns and noun phrase structures. This nominal style contrasts with the structure of, for example, conversation, which relies on the use of more verbs and clausal structures (see Biber 1988; Biber & Gray 2010, 2016). Halliday's work on scientific writing has focused on describing 'grammatical metaphor' (see Halliday 2004 for a collection of key works on nominalization and grammatical metaphor in science writing), whereby processes and actions typically expressed with verbs are nominalized, that is, verbs are changed to nouns through a process of grammatical metaphor.

Much of Biber's work has empirically documented the nominal style of academic writing using large-scale corpus analyses that compare academic writing to other general registers like conversation, fiction, and newspaper writing. In a multi-dimensional analysis using factor analysis to identify how 67 linguistic features co-occur on a statistical basis, Biber (1988) shows that academic writing commonly uses features associated with an 'informational' purpose, such as nouns, prepositions, and attributive adjectives, while relying much less on linguistic features associated with the 'involved' nature of conversation such as private

verbs, *that*-deletions, personal pronouns, WH-questions, modals, and WH-clauses among others (see Biber 1988: Chapter 6).[1]

In perhaps the most comprehensive descriptive reference grammar to date, Biber et al. (1999) describe the distributions of a full range of lexical and grammatical structures in English in *The Longman Grammar of Spoken and Written English (LGSWE)*, comparing academic writing, conversation, newspaper writing, and fiction. In the *LGSWE*, academic prose is represented by extracts from academic books and research articles in 13 disciplines ranging from agriculture to computing to mathematics to sociology (see Biber et al. 1999: 32–33). Although a full summary of the characteristics of academic prose is not possible here, Biber (2006: Chapter 1) provides a comprehensive summary of grammatical features that occur particularly frequently in academic prose based on the *LGSWE*. In particular, Biber (2006: 15–18) notes quite a few grammatical features associated with noun phrases that are either most common or very common in academic prose when compared with other registers, including the overall use of nouns, nouns as pre-modifiers, nominalizations, adjectives, prepositions, *of*-phrases, relative clauses with *which*, and noun post-modifying participle clauses.

More recently, Biber and colleagues have focused on expanding this analysis of features which contribute to the nominal style of academic writing, both synchronically and diachronically (see Biber & Gray 2010, 2016; Biber, Gray & Poonpon 2011; Biber et al. 2011). In particular, this research has documented the prevalence of nouns and phrasal modifiers in academic writing, such as attributive adjectives, nouns as nominal premodifiers, and prepositional phrases as post-modifiers. Consider the following two examples from a quantitative biology article and a quantitative research article in applied linguistics (head nouns with pre- or post-modification are **bolded**, phrasal post-modifiers are underlined, and adjectives and nouns as nominal premodifiers are *italicized)*:

1.1 Given their **importance** in the **functioning** of *arid* and *semiarid* **ecosystems,** restoring these crusts may contribute to the **recovery** of *ecosystem* **functionality** in *degraded* **areas**. [BIO-QT-12]

1.2 The *main* **aim** of conducting this study was to investigate the *foreign language learning* **needs**, **wants** and **desires** of *undergraduate medical sciences* **students** studying in **faculties** of nursing and midwifery in *various* **universities** in Iran. [AL-QT-10]

1. In fact, 'Dimension 1' of the 1988 MD analysis showed a clear cline of variation that distinguished between written and spoken registers generally. However, academic prose had one of the lowest dimensions scores, indicating its high use of informational features and low use of interactional features.

These two sentences illustrate the dense embedding of nominal modifiers within noun phrases, resulting in condensed structures in which a head noun is modified, often multiple times, in order to pack a great deal of information into a few noun phrases. This style is primarily restricted to academic prose, and Biber and colleagues connect this nominal style to the informational purpose and highly specialized readership of academic prose.

While the research briefly summarized above has focused on describing academic writing without direct consideration of disciplinary differences, a great deal of research has also concentrated on describing the linguistic characteristics of writing in the disciplines, summarized in the next section.

1.3 Linguistic variation and disciplinary writing

Across language-related fields, increasing attention has been being paid to language use in relation to specific disciplinary practices. Movements such as Writing across the Curriculum (WAC; see Russell 1991 for a comprehensive history) have brought attention to the importance of teaching writing within specific curricular areas and disciplines, rather than independent of specific content areas and discourse communities. Alongside the rise of WAC, awareness that language use varies in meaningful ways across disciplines has also grown. Because of this awareness, studies investigating disciplinary language use have flourished, and a look at any journal focusing on English for Academic Purposes (EAP) will reveal a large body of research about language use in the disciplines from a variety of perspectives and research methodologies.

One approach to studying the linguistic characteristics of writing in specific disciplines is to focus on one register in one discipline – to provide detailed descriptions of a particular type of text within the context of a specific disciplinary community. However, the majority of studies concerned with discipline-specific language take a comparative approach to describing the linguistic characteristics of disciplines and/or registers. These studies can be categorized into two major types: (1) those that compare two registers within a discipline or disciplines, and (2) those that compare the same register in multiple disciplines.

As with the treatment of discipline in academic writing research, the registers under investigation also vary in level of specificity. Some studies group several registers together to represent academic language use. For example, Fuertes-Olivera (2007) examines lexical gender in a variety of registers in business, including research articles, product descriptions, political speeches, and governmental reports. Others compare and contrast two or more registers as Koutsantoni (2006) does in an investigation of hedges in research articles and student theses. A great

majority of investigations focus on one particular register. Although research articles are by far the most widely-studied register, other written academic registers that have been studied include textbooks (Biber, Conrad & Cortes 2004; Conrad 1996a; Freddi 2005; Moore 2002), Ph.D. or master's theses/dissertations (Bunton 2005; Charles 2006a, 2006b; Samraj 2008), peer review reports (Fortanet 2008), 'comment' articles (Lewin 2005), book reviews (Groom 2005), and business reports (Yeung 2007).

Many studies also explore the use of linguistic features within specific sections of the target registers. For example, Chen and Ge (2007) examine the distributions of Academic Word List (AWL) words in different sections of research articles. Martínez (2003, 2005) investigates thematic structure and 1st person pronouns in a corpus of biology articles, comparing uses across different sections in the articles. Samraj (2005) looks at the rhetorical structure of research article abstracts and introductions.

While a great deal of the research on academic writing focuses on describing linguistic phenomenon within a genre or single discipline, a smaller body of literature investigates differences in similar genres across disciplines. Since the purpose of this book is to describe disciplinary differences, I will explore these studies in a bit more detail and examine the varying types of linguistic structures that are focused on in these studies.

Aspects of lexis are one feature that is investigated in academic writing, including lexical bundles, keyword analysis, and collocational analysis. Cortes (2004) examines four-word lexical bundles in research articles from history and biology, finding that history articles employed bundles out of the two structural groups noun phrases and prepositional phrases while biology articles employed a much wider range of structures. Cortes also finds that despite using different lexical bundles, both disciplines employed bundles for similar functions.

In her analysis of keywords in introduction chapters in applied linguistics textbooks, Freddi (2005) finds that when compared to a reference corpus, applied linguistics textbook introductions use words that represent processes, logical connections, and interpersonal relationships. Freddi interprets this to be a way that the authors create relationships with the readers of the textbooks and position themselves as a teacher or researcher and those reading the textbook as students.

A second type of analysis investigates grammatical structures. Groom (2005) is an example of such a study, investigating extraposed adjective-controlled *to*- and *that*-clauses in research articles and book reviews in history and literary criticism. Groom finds that the phraseological patterns of these structures differ across the two genres and disciplines. In fact, *that* complement clauses are an often-studied grammatical structure in academic prose (Parkinson 2013; Charles 2006a, 2006b, 2007; Hyland & Tse 2005). *That* complement clauses are productive structures in

academic prose, as they are integral to citing others' work and presenting claims, and are often indicators of an author's stance. Charles (2006b) examines how writers of theses in politics and materials science use verb-controlled *that* clauses to report information, considering the subjects of the clauses (including *it*) and the use of passive voice. Charles's study illustrates a key trend in the investigation of grammatical constructs in academic prose: when investigating grammatical features, researchers typically consider many aspects of the linguistic context and make connections to the specific purposes that the grammatical structures fulfill in academic prose.

In fact, relatively few studies examine single grammatical features. Rather, researchers focus on a collection of lexical and grammatical features that work together to create some type of functional result. For example, stance (as marked by a collection of lexical and grammatical markers) is a widely-researched topic in academic writing. Hyland (1998) compares stance markers in eight disciplines, finding that disciplines in the humanities/social sciences exhibited nearly 2.5 times as many stance markers than the sciences.

Afros and Schryer (2009) also use a combination of lexical and grammatical features to investigate a single construct: self-promotion through the use of lexical items, coordination, comment clauses, personal pronouns, and lexical cohesion. Afros and Schryer found that literary scholars relied on intensifiers to persuade readers to believe their claims, while linguistics scholars relied on self-citation.

Hyland (2002a) and Swales, Ahmad, Chang, Chavez, Dressen, and Seymour (1998) study the use of commands (termed directives and imperatives respectively) in academic research articles. While Swales et al. identify commands based on the criteria that a main verb or emphatic *do* occurs in the base form with no modals and no surface subject, Hyland identifies commands based on a set of 80 lexical search terms (it should be noted that Hyland uses the results from Swales et al. to choose his search terms). Both Hyland and Swales et al. find discipline-specific differences in the use of commands. For example, Hyland finds that 'hard' disciplines (such as mechanical engineering and physics) used many more directives that were intended to direct the reader through procedures and conclusions than the 'soft' disciplines (such as philosophy). Swales et al. find that fields relying on mathematical reasoning employ more commands. These two studies illustrate that, to this point, much of the research that has investigated more than one or two disciplines has interpreted results in terms of categories of disciplines, such as 'hard' or 'soft' disciplines, and disciplinary variation seems to follow generally along those lines.

A final thread found in research on disciplinary writing examines the organizational structure of genres, many times by employing a move analysis. Because of the primarily qualitative nature of move analysis, many studies focus on only

one discipline (e.g., Nwogu's 1997 description of medical research articles), and others investigate only a specific section of the texts. For example, Ozturk (2007) uses Swales's (1990) CARS model to describe introductions in research articles in applied linguistics, while Basturkmen (2012) analyzes discussion sections of dentistry research articles.

However, a few of these move analysis studies do compare disciplines. For example, Stoller and Robinson (2013) compare the rhetorical structure of chemistry and biochemistry articles. Samraj (2005) conducts a move analysis of abstracts and introductions in conservation biology and wildlife behavior, finding that while abstracts and introductions in conservation biology share similar functions and organizations, the two genres in wildlife behavior are not as similar. Holmes (1997) compares discussion sections in articles in history, political science, and sociology. He finds that these three social sciences have similarities and differences when compared to previous findings about discussion sections in natural science research articles, and that the three social sciences disciplines varied amongst themselves as well. Holmes concludes, however, that the political science and sociology discussion sections were sufficiently similar to natural science discussion sections to call them the same genre, with history texts showing much more variation (such as being brief and not containing a cyclical structure). Because of the qualitative nature of move analysis, it has not typically been applied to large-scale corpus-based studies until a recent volume by Biber, Connor, and Upton (2007) and the dissertation by Kanoksilapatham (2005b). However, corpus-based research on move structure is on the rise, and researchers are increasingly paying attention to specific lexical and grammatical features as they are associated with specific rhetorical moves (e.g., Cortes 2013 on lexical bundles associated with moves in RA introductions).

As this summary of research shows, research articles are perhaps the most commonly researched register within academic writing (although there has also been a great deal of research into general academic writing by L2 speakers, as well as novice L1 writers. For examples, see Parkinson & Musgrave 2014; Callies 2013; Grant & Ginther 2000; Green, Christopher & Mei 2000; Hardy & Römer 2013; Hinkel 2003; Jarvis, Grant, Bikowski & Ferris 2003; Schleppegrell 1996; Spycher 2007; Altenberg & Granger 2001; Archer 2008; Flowerdew 2006; Loudermilk 2007). Table 1.1 below summarizes studies on academic research articles, along with the linguistic features investigated in the studies.[2] Table 1.1 illustrates

2. Because the current study focuses on the distribution of lexical and grammatical features of research articles, I have excluded the large body of research using a Swales-inspired moves analysis from this summary table. Interested readers, however, can see the following

that a wide variety of linguistic features are investigated in these studies, and research articles are often compared to either other registers within the realm of academic writing, to novice writer academic writing, or to other general registers like conversation.

As can be seen in the brief literature review above and in Table 1.1, academic writing has been widely researched. In this broad research base, research articles are often compared to student-produced registers (e.g., theses and dissertations written by advanced graduate students, 2nd language writers at a variety of levels), as well as to other sub-registers within academic writing such as textbooks and academic lectures (and less frequently to registers like book reviews and editorials). About as often, however, research articles are not compared to other registers at all.

In terms of the linguistic features that are focused on this research, Table 1.1 documents that research articles are often analyzed for their use of functionally-related lexical and grammatical features. This is illustrated by the fact that much of the terminology used to describe the linguistic features of interest represents functional constructs rather than specific lexical or grammatical structures. For example, a great deal of research has focused on the ways in which authors encode values and judgments in their texts under terms such as hedging, stance, boosting, appraisal, and evaluation. Other examples that illustrate this focus on functional constructs include metadiscourse, citation, self-mention, argument structure, naming conventions, new knowledge claims, and so on.

1.4 Trends and gaps in the study of disciplinary writing

The literature review presented in Section 1.3 illustrates the diverse nature of linguistic studies of academic writing, and of academic research articles more specifically. Despite the large foundation of research into disciplinary writing, it is difficult to arrive at a comprehensive description of disciplinary variation. More specifically, the following trends emerge from this review. First, most large-scale investigations that consider a wide range of linguistic features focus on comparing academic prose in general with other broadly-defined registers like conversation or news. While these comparisons are inherently interesting and useful, providing

---

for moves analyses of research articles: Koutsantoni (2006), Basturkmen (2009), Bhatia (1997), Brett (1994), Bruce (2008, 2009), Bunton (2005), Holmes (1997), Kanoksilapatham (2005a,b), Lim (2006), Lin & Evans (2012); Ozturk (2007), Ruiying & Allison (2004), Samraj (2002, 2004, 2005), Stoller & Robinson (2013).

Table 1.1. Summary of studies investigating language use in academic research articles

Study	Disciplines	Registers Compared to Research Articles	Linguistic Features
Afros & Schryer (2009)	language and literary studies	–	promotional metadiscourse (moves, lexico-grammatical markers)
Aktas & Cortes (2008)	art & design, computer science, economics, environmental engineering, physics, astronomy	L2 graduate-level academic writing	shell nouns
Biber & Gray (2010)	biology, medicine, ecology, physiology, education, psychology, history	conversation	grammatical features of complexity and elaboration
Biber & Gray (2013)	science vs. humanities	–	grammatical features of complexity and elaboration
Biber, Csomay, Jones & Keck (2004)	various	academic lectures, university textbooks	vocabulary-based discourse units
Chen & Ge (2007)	medicine	–	AWL word families
Conrad (1996b)	ecology	composition textbooks, ecology textbooks	various
Cortes (2013)	13 disciplines	RA introductions	associating lexical bundles with discourse moves
Dahl (2008)	economics, linguistics	–	new knowledge claims
Diani (2008)	linguistics, history, economics	academic lectures, book reviews	emphasizer *really*
Dueñas (2007)	business management	–	self-mentions and citations

(*Continued*)

Table 1.1. (Continued) Summary of studies investigating language use in academic research articles

Study	Disciplines	Registers Compared to Research Articles	Linguistic Features
Fang, Schleppegrell & Cox (2006)	various	various	nouns
Feng & Hu (2014)	applied linguistics, education, psychology	quantitative vs. qualitative RAs (post-method sections)	metadiscourse
Gillaerts & Van de Velde (2010)	applied linguistics (pragmatics)	**RA abstracts only	interactional metadiscourse (hedges, boosters, attitude markers)
Gosden (1992)	science	–	marked themes
Gray (2010)	education, sociology	–	demonstrative pronouns and determiners; shell nouns
Groom (2005)	history, literary criticism	book reviews	it + V-ling + ADJ + that/to
Harwood (2005a)	computer science	student writing (project reports, MA theses)	first person pronouns
Harwood (2005b)	business management, computing science, economics, physics	–	first person pronouns
Hemais (2001)	marketing	–	sentence subjects, citations, reporting verbs
Hewings & Hewings (2002)	business	MA dissertations	anticipatory it, extraposed subjects
Hewings, Lillis & Vladimirou (2010)	psychology	–	citations
Hyland & Tse (2007)	sciences, engineering, social sciences	textbooks, book reviews, letters, MA & Ph.D. theses, student projects	academic vocabulary (AWL)
Hyland (1996)	molecular biology	–	hedging
Hyland (1999a)	microbiology, marketing, applied linguistics	textbooks	metadiscourse

(*Continued*)

Table 1.1. (Continued)

Study	Disciplines	Registers Compared to Research Articles	Linguistic Features
Hyland (1999b)	philosophy, sociology, applied linguistics, physics, electrical engineering, marketing, mechanical engineering, biology	–	citation & attribution
Hyland (2001a)	8 disciplines (see Hyland 1999b)	–	2nd person pronouns, inclusive pronouns, questions, directives, etc.
Hyland (2001b)	8 disciplines (ibid.)	–	self-mention
Hyland (2002a)	8 disciplines (ibid.)	–	directives
Hyland (2002b)	8 disciplines (ibid.)	L2 academic writing	authorial identity
Hyland (2007)	8 disciplines (ibid.)	–	exemplifying and reformulating (elaboration)
Hyland (2008)	electrical engineering, biology, business, applied linguistics	MA & Ph.D. theses	lexical bundles
Hyland (2010)	science and engineering	popular science articles	proximity, argument structure
Khedri, Heng, & Ebrahimi (2013)	applied linguistics, economics	research article abstracts only	metadiscourse
Koutsantoni (2004)	electrical engineering	–	appraisal
Koutsantoni (2006)	electrical engineering	research theses	stance
Kuo (1999)	science	–	personal pronouns
Kwan (2012)	2 areas within Information Systems	–	strategies for making evaluations within specific moves in the CARS model
Lee & Chen (2009)	linguistics/applied linguistics	dissertations, student written assignments	keywords
Marco (2000)	medicine	–	collocational frameworks
Martínez (2003)	physical science, biological science, social science	book chapters	finite clauses, transitivity structures

(*Continued*)

Table 1.1. (Continued) Summary of studies investigating language use in academic research articles

Study	Disciplines	Registers Compared to Research Articles	Linguistic Features
Martínez (2005)	biology	NNES manuscripts	first person pronouns
Martínez, Beck & Panza (2009)	agriculture	–	vocabulary
Norman (2003)	bio-medical	RA abstracts only	naming conventions
Parkinson (2013)	social sciences	across sections of RAs	*that*-complement clauses
Peacock (2006)	business, language and linguistics, physics, administration, law, environmental science	–	boosting
Salager-Meyer (1994)	medicine	–	hedges
Stotesbury (2003)	humanities, social sciences, natural sciences	RA abstracts only	evaluation
Swales et al. (1998)	philosophy, sociology, applied linguistics, physics, electrical engineering, marketing, mechanical engineering, biology	–	imperatives
Tarone, Dwyer, Gillette, Icke (1998)	astrophysics	–	passive and active voice
Thomas & Hawes (1994)	medicine	–	reporting verbs
Tucker (2003)	art history	–	evaluation
Vongpumivitch, Huang & Change (2009)	applied linguistics	–	use of AWL words
Warchal (2010)	linguistics	–	conditional clauses
Webber (1994)	medicine	editorials, letters	questions
Williams (1996)	medicine	–	lexical verb use

much-needed accounts of variation in language use broadly, they do not consider disciplinary differences in their research design. As a result, these studies cannot inform discipline-specific language instruction or help us explore the inner workings of disciplinary cultures.

Second, smaller-scale investigations into single disciplines are common. While useful for describing particular disciplines, investigations of single disciplines provide little contrastive information about how the features of interest might vary across disciplines. Also common are studies comparing a small number of disciplines; the comparative approach taken in these studies allows for more contrastive information about how disciplines might vary. However, with both of these lines of inquiry, it's difficult to arrive at a comprehensive picture of disciplinary variation, as replication studies that employ directly comparable methodologies in different disciplines are carried out relatively infrequently.

Third, while an abundance of literature describes research articles, two main problems arise from this trend. Giving preference to research articles above other types of articles published in academic journals ignores many other modes of the transmission of knowledge in academic disciplines and limits our knowledge of the discourse practices within disciplines. This is an important consideration, and one that will need to be further addressed in linguistic research of academic writing in the future, particularly as these studies of research articles often serve the basis for corpus-informed writing instruction.

For the purposes of the present book, a second problem that arises from this focus on research articles is of interest: research articles are generally not a finely-defined register in that similarities across articles which report on distinct research methodologies and disciplines are assumed but not often investigated. In reality, the differing nature of research may be leading to substantial linguistic variation not being captured by the studies that have been conducted to date.

A limited body of research has acknowledged the presence of registers other than research articles in academic journals: Magnet & Carnet (2006), Flowerdew & Dudley-Evans (2002), and Giannoni (2008) on editorials/letters to the editors in academic journals; Fortanet (2008) on evaluative language in peer review referee reports in applied linguistics and business. However, these studies consider types of publications in journals that are not primarily research reports; very little research considers differences in *types* of research reports. Instead, analyses of research reports generally consider all research reports to be a single register, declining to distinguish between articles which report on, for example, qualitative versus quantitative research, or case study research versus survey research, and so on.

Despite the lack of research investigating the linguistic differences in these areas, there is an implicit recognition that such differences exist, and that novice writers may struggle with the characteristics of a particular register. For example, Belcher and Hirvela (2005) focus on the writers of qualitative doctoral dissertations, looking at their motivations and commitments to what is seen as a challenging register for L2 writers, and implicitly acknowledging that

writing up qualitative research is inherently different than writing up quantitative research. Yet, there has been little systematic inquiry into these differences, leading to a lack of empirical evidence of the linguistic characteristics of these various registers.

A few studies have acknowledged different types of research articles, but have primarily used these distinctions to restrict the analysis that is undertaken, or to inform corpus design. For example, Williams (1996) analyzes the use of lexical verbs in clinical and experimental medical research articles. Despite a general claim that descriptions of different types of medical reports are needed for comprehensively informing ESP materials, Williams does not offer explanations of how the two types of reports differ in terms of their non-linguistic characteristics. Likewise, Vande Kopple (1994) restricts his study of complex subjects to experimental science articles, acknowledging that the same patterns may not be found in theoretical science articles (but does not carry out an empirical comparison to test that hypothesis).

There are two recent exceptions to this trend, and the results of both studies indicate that making distinctions between types of research reports is an area where more research is warranted. Cao and Hu (2014) compare "post-method" sections of research articles in three disciplines: applied linguistics, education, and psychology, directly contrasting quantitative and qualitative research articles in the three disciplines. Cao and Hu (2014: 16) find that the metadiscourse markers employed in the research articles in their corpora vary in ways that can be associated with both disciplinary factors, as well as with "the contrasting epistemologies underlying the qualitative and quantitative research paradigms".

Kwan, Chan, and Lam (2012) also suggest that research paradigm correlates to differences in the discourse style of published academic research articles. Taking a slightly different approach, Kwan et al. contrast two sub-disciplines within the field of Information Systems, comparing journals which publish research following a behavioral science research paradigm with a design science research paradigm. In their analysis of the rhetorical strategies that authors use to evaluate previous research while carrying out moves from Swales' CARS model, Kwan et al. find that authors of research in the two paradigms rely on different types of evaluation strategies. They interpret these patterns as reflecting the epistemological orientations of the two different paradigms. In summary, while few studies have considered research method or epistemologies as influences on language variation in research article, what little research exists has uncovered notable patterns in the language and discourse styles.

A final comment that can be made based on the summary of diverse studies above in Table 1.1 is that the research agenda into academic research articles has been rather piecemeal, or has tended to focus on a few narrow areas

(e.g., metadiscourse, stance, moves or rhetorical structure). It is difficult to synthesize large-scale findings on disciplinary variation based on this research, and little research has focused on whether there is variation in the core lexical and grammatical structure of writing across disciplines. In other words, while this research has provided valuable and detailed insights into particular registers and particular disciplines, it is difficult to get an overall picture of the wide range of linguistic variation that is occurring in research articles in different academic disciplines. Thus, the goal of the present book is to take a wide variety of core linguistic features and analyze them systematically across 6 disciplines (philosophy, history, political science, applied linguistics, biology, and physics). To address the three major gaps presented here, I first develop a framework for identifying different academic journal registers, including different types of research articles, across many disciplines. I then carry out a large-scale linguistic analysis to compare and contrast research articles within and across disciplines, creating rich descriptions of a variety of registers and disciplines within academic journal writing.

1.5 Overview of the book: Applying corpus analytical approaches to disciplinary register variation

Corpus linguistics methodologies are useful for investigating differences across disciplines and registers, in part because of the relative ease with which comparisons across varieties can be made using the quantitative data that result from corpus analysis. The research undertaken in this book aims to reveal patterns of linguistic variation within and across disciplines through a consideration of the varied nature of research articles in six disciplines. However, before any corpus-based linguistic analyses can be carried out, work has to be done to identify the range of possible journal registers, select disciplines to be included in the analysis, and determine criteria by which texts can be reliably classified into sub-corpora representing the various registers. That is, non-linguistic analyses must be carried out to inform the design of a corpus that can be used to reliably represent the varieties of interest. This type of non-linguistic analysis is a crucial stage in addressing the issue of external or situational representativeness (McEnery et al. 2006; Biber 1993) in corpus design: the extent to which a corpus is designed to represent "the range of text types in the target population" (Biber 1993: 243).

Furthermore, the registers under investigation need to be described in detail for their purposes, topics, authorship, and so on in order to help explain the patterns of language use that are uncovered during the analysis stage. All of these tasks can be accomplished by undertaking analyses of the non-linguistic, or

situational characteristics, of (a) the target registers at the corpus design stage, and (b) the texts in the resulting corpus. While the situational analysis of the target register can inform the corpus design and increase the probability that external representativeness can be achieved, a situational analysis of the texts in the resulting corpus enables the empirical evaluation of that representativeness, and is needed for interpreting the patterns of variation that are discovered in that corpus.

As a result, the book includes two main types of analyses: (1) analyses of the non-linguistic characteristics of the broad domain of academic journal writing, and of the texts in the corpus more specifically; and (2) linguistic analyses of the patterns of variation in the corpora. Figure 1.1 visually represents the various components of the research being reported in the present book, illustrating the relationship between the individual analyses and the sources of information for these analyses. In the figure, squares represent non-linguistic analyses or sources of information, with sources differentiated from analyses with a dashed outline. Ovals indicate the linguistic analyses carried out in the study. Dashed arrows indicate where information from an analysis/source has informed another component of the study in some way.

Figure 1.1 shows that the study is divided between analyses and information that comes from the world external to the corpus, and analyses that are based on the corpus collected for the study. The analyses of and sources from the broader context outside the corpus all inform the design of the corpus in some way. The main analysis here is a survey of 11 disciplines that identifies ten journal registers and documents how frequently the 11 disciplines publish those registers. This survey, detailed in Chapter 2, is used to inform the choice of registers and disciplines for the Academic Journal Register Corpus, as well as to form the basis for the creation of operational definitions that can be applied to classify texts into the journal registers. However, this survey is not the only criteria in determining the corpus design; meetings with experts within the fields of interest play an important role in refining the general operational definitions for use during corpus collection.

The placement of the circle encompassing corpus design is symbolic of the bridging role that a corpus plays in linguistic research. While the corpus is the basis for the linguistic analyses, it is also intended as a representation of the larger world outside the corpus. That is, it is intended to reflect a type of language that exists in the larger context.

Once the corpus is constructed based on input from the various external analyses and sources, corpus-based research can take place. Going back to Figure 1.1, we can see that the study encompasses one corpus-based situational analysis. The corpus-based situational analysis draws on the analyses and sources from the wider situational context to identify specific features that can be used to characterize the texts (and in turn the registers) in the corpus. Two analyses of the situational

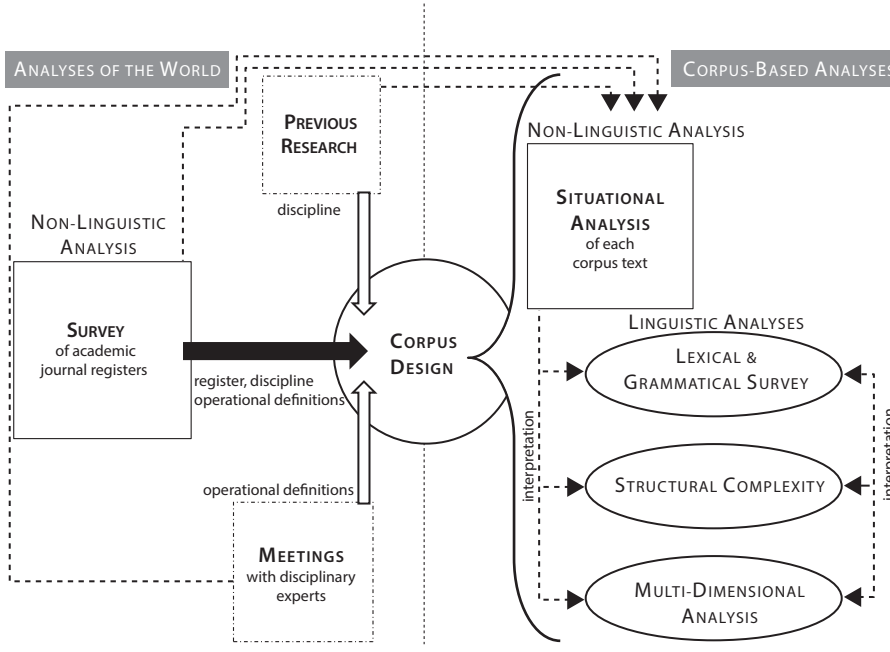

Figure 1.1. Overview of the study: Relationships between and among linguistic and non-linguistic analyses

characteristics of the registers take very different approaches. In the first approach (the survey reported in Chapter 2), the purpose is to describe the broad domain of academic journal publishing with the intent of informing the corpus design and understanding how the registers differ. This approach uses a survey method that focuses on the holistic context of journal publishing. In the second approach, the purpose is to describe the non-linguistic characteristics of the corpus texts themselves. The second approach accordingly employs a corpus-based method in which a framework for describing these non-linguistic (or situational) characteristics is applied to each text in the corpus. It is important to note that these two analyses are not completely separate; rather, information gained during the survey of the domain of academic journal writing informs the development of the framework used in the corpus-based situational analysis of the corpus texts (described in Chapter 4).The situational analyses reported in this book are conducted systematically and empirically, and serve as the foundation for the corpus compilation and linguistic analyses. Many register-based studies, and studies based on academic writing more specifically, begin with pre-conceived, broad definitions of the registers of interest. Some studies survey the field to inform corpus composition, while others analyze certain aspects of the corpus after compilation is complete (or even

after analyses are completed, as a way of interpreting the linguistic results). However, very few studies conduct both types of analyses. Moreover, situational analyses are often applied on the 'register' level. That is, a target register is described in terms of its typical characteristics, with little attention being paid to the individual texts and what their non-linguistic characteristics are. This typically results in registers being described fairly broadly (e.g., informational purpose of academic writing vs. interpersonal relationship management for conversation).

Thus, the use of two situational analyses, one to inform corpus design and one to describe the corpus composition, distinguishes the research reported on in the present book. On one hand, these situational analyses and sources of information help ensure and document that the corpus has 'external' or 'situational' representativeness – that the corpus texts represent the types of texts found in the target domain (see McEnery et al. 2006; Biber 1993). On the other hand, the situational analyses provide crucial information needed in order to interpret the patterns of linguistic variation that are subsequently uncovered in the corpus.

Figure 1.1 illustrates this relationship between situational analyses and corpus design and linguistic analysis. The rich description of the texts that results from the situational analysis is applied within the linguistic analyses to help interpret and explain the uses and distributions of the linguistic devices of interest, as indicated by the dashed arrows in Figure 1.3. Three linguistic analyses are carried out in Chapters 5 through 7.

Chapter 5 provides a descriptive overview of the general grammatical characteristics of academic writing across disciplinary journal registers. The perspective is strictly lexical and grammatical, focusing on the distribution of different word classes and the most commonly occurring classes of lexical items. Chapter 6 moves on to an examination of structural complexity, with features being linked to functions of compressing and elaborating information (based on Biber & Gray 2010; Biber, Gray & Poonpon 2011). This analysis is lexico-grammatical in nature, as the units being analyzed are identified based on a combination of grammatical patterns and frequent lexical items that occur in those grammatical contexts.

Chapter 7 presents a comprehensive register description. Using the multidimensional analysis methodology developed by Biber (1988, 1995), this study considers the co-occurrence patterns of 70 lexical and grammatical features through the statistical technique of exploratory factor analysis. Once the patterns in the use of linguistic features have been determined statistically, those patterns are interpreted according to the functions that those groupings of linguistic features carry out. Biber (2010) describes MD studies as uncovering 'dimensions' of variation as previously-unrecognized linguistic constructs that emerge from the inductive analysis of quantitative patterns in the corpus (see Biber 2010: 179).

In addition to using the situational analyses to interpret the quantitative trends found in the linguistic analyses, the findings from the linguistic analyses themselves can complement one another. In particular, the more narrowly-focused analyses on specific features in Chapters 5 and 6 are particularly influential in interpreting the co-occurrence patterns of those same features as they are uncovered in the multi-dimensional analysis (Chapter 7). Chapter 8 then brings together the results of both the situational and linguistic analyses, to synthesize the major patterns of disciplinary and register variation, and to discuss the implications of these findings for research and disciplinary variation in academic writing.