

More than discipline: uncovering multi-dimensional patterns of variation in academic research articles

Bethany Gray¹

Abstract

Research analysing linguistic variation in research articles (RAs) across academic disciplines typically employs a relatively coarse definition of RAs: any text published in an academic journal that reports on primary research. This definition assumes that articles within a single discipline which report on distinct research methodologies are similar linguistically. Yet little empirical work has investigated this assumption. This study uses multi-dimensional analysis to analyse variation in the use of seventy lexical and grammatical features in 270 research articles representing three sub-registers (theoretical, qualitative and quantitative research reports) in six disciplines (philosophy, history, applied linguistics, political science, biology and physics). The resulting dimensions of variation indicate that linguistic variation occurs along multiple parameters, not only across disciplinary lines. For example, variation also corresponds to the differing purposes and types of evidence associated with the three research paradigms. This article explores each of these complex patterns of variation to move towards a more comprehensive understanding of language use within and across disciplines, and considers the implications of these findings for future corpus-based analyses of disciplinary variation.

Keywords: academic writing, disciplinary variation, MD analysis, nominal style

1. Introduction

Research analysing linguistic variation across academic disciplines has often focussed on the research article (RA) as the primary professional written

¹ Department of English, Iowa State University, 206 Ross Hall, Ames, IA 50011, USA.
Correspondence to: Bethany Gray, e-mail: begray@iastate.edu

academic register. This research has been founded on the belief that published texts represent the knowledge held by a disciplinary community (Bazerman, 1994), and that disciplines utilise linguistic resources in varied ways to construct these texts. Simply put, language use varies across disciplines. In this study, I extend our knowledge of such disciplinary variation using multi-dimensional (MD) analysis (Biber, 1988) to investigate sub-registers of published research articles in six disciplines.

1.1 The research article and disciplinary variation

Previous research on the linguistic characteristics of RAs across disciplines has taken several approaches. One approach has been to focus on RAs in a single discipline, offering a detailed analysis of a particular lexical, grammatical or rhetorical device (e.g., Afros and Schryer, 2009; Hemaïs, 2001; Hyland, 1996; and Warchal, 2010). A second approach has been to compare the frequencies and functions of linguistic features in multiple academic registers within a single discipline (or a small number of disciplines). While comparisons between novice and/or L2 writer production and published RAs are common (e.g., Harwood, 2005a; Hewings and Hewings, 2002; Koutsantoni, 2006; and Martínez, 2005), they have also been compared to textbooks (e.g., Biber *et al.*, 2002; and Hyland, 1999), book reviews (e.g., Diani, 2008; Groom, 2005; and Hyland and Tse, 2007), editorials (e.g., Webber, 1994) and popular science texts (Hyland, 2010).

Perhaps the most widely adopted approach, however, has been to consider two or more disciplines, with a focus on a single register: the research article (e.g., Harwood, 2005b; Peacock, 2006; and Hyland's extensive body of work on RAs in eight disciplines). All three approaches have contributed to our understanding of disciplinary discourse, and the brief references included here illustrate that attention has been given to a wide range of linguistic features, from lexical and grammatical features, to the rhetorical structure of texts, and to compilations of linguistic features with shared functions (e.g., markers of stance).

Within this canon, research articles are, in general, coarsely defined to include any text published in an academic journal that reports on a form of primary research. Relatively little attention is paid to the nature of those research activities. For example, in cross-disciplinary research, interpretations of linguistic differences often involve commentary on the differing nature of disciplines. Yet texts are still grouped under a relatively general register of 'academic research articles'. Consequently, a philosophy article arguing Kant's theory of perception is considered to belong to the same register as a physics article on photon absorption—despite differences in their characteristics, such as research type (theoretical *versus* experimental) and the type of evidence relied upon (logical and cognitively based arguments *versus* observed data).

Likewise, single-discipline research often forgoes considerations of intra-discipline variation in types of research reports, and instead focusses on comprehensive descriptions of the use of linguistic features across relatively broad register distinctions (such as student writing, textbooks and book reviews) or within the general 'RA' register. For example, no distinction is typically made between an applied linguistics article documenting an ethnographic observation of a language classroom and one that reports an experimental study of explicit feedback and question formation in a second language.

In spite of this practice, there is at least implicit evidence that characteristics such as research paradigm play an important role in how academics write up their research. Belcher and Hirvela (2005) look at graduate student motivations in writing dissertations based on qualitative research; the qualitative dissertation is seen as an especially challenging register for L2 writers, acknowledging that it is inherently different from other types of research reports. Likewise, Williams (1996) makes a point of including both clinical and experimental medical RAs in his dataset, claiming that analyses of both are needed for informing ESP materials and implying that the language of these texts differs in some way.

Despite wide recognition that academic disciplines utilise distinct research paradigms, corpus-based studies of disciplinary writing have not systematically accounted for such differences in the design of corpora. The result is that disciplinary variation could be being confounded with differences in the type of research that disciplines employ. Likewise, we know little about intra-discipline variation that may be associated with the variety of research paradigms that are valued within a discipline.

The purpose of this study is to take into account these varied realisations of research reports while analysing the overall patterns of linguistic variation in RAs in fundamentally diverse disciplines. To do so, I use a multi-dimensional analysis (Biber, 1988) to identify patterns of linguistic co-occurrence for a wide range of linguistic features. This research seeks to build upon previous research that has described the structure of written academic language (described in Section 1.2), to identify specialised patterns of variation that correspond to sub-registers within the general register of academic research articles. Section 2 focusses on the corpus of RA sub-registers used in this study, while Section 3 summarises the multi-dimensional analysis procedures. In Section 4, the four dimensions of variation revealed by the analysis are described and illustrated.

1.2 The linguistic characteristics of academic language: motivating a new multi-dimensional analysis of research articles

Linguists have long worked towards describing the characteristics of academic prose that differentiate it from other spoken and written registers. Much of this research has focussed on the nominal style of academic writing,

or the heavy reliance on nouns and associated noun-phrase structures (e.g., nouns, nominalisations, nouns as noun modifiers, adjectives, prepositional phrases as noun modifiers, and participle clauses modifying nouns; see Banks, 2005; Biber, 1988; Biber *et al.*, 1999; Biber and Gray, 2010; Biber *et al.*, 2011; Fang *et al.*, 2006; Halliday, 2004; and Wells, 1960). Recent diachronic studies have shown that science writing has adopted this dense nominal style to a much greater extent than social science or humanities writing over the past 100 years (see Biber and Gray, 2013). This finding supports the premise that disciplinary variation in the use of core grammatical features exists, although this line of research has generally not been concerned with disciplinary variation.

In contrast, most research on disciplinary variation has provided detailed analyses of specific lexical and grammatical features. While these studies help us to understand particular features and functions of language, it is difficult to synthesise findings into a comprehensive description of linguistic variation, especially one which can factor in issues such as the nature of research within and across disciplines. Such a comprehensive description of variation can be achieved through multi-dimensional (MD) analysis. This method is based on the recognition that:

- (1) Language variation occurs across multiple parameters, and these parameters correspond to different sets of co-occurring features;
- (2) Variation is realised along a continuum of relative frequency of use, rather than dichotomies of use or non-use; and,
- (3) Such 'dimensions' of variation can be identified based on statistical co-occurrence patterns of a comprehensive set of linguistic features using factor analysis.

(Conrad and Biber, 2001)

2. Representing research articles: a corpus of disciplinary writing in three registers

The first step in building a principled RA corpus that also represents distinct types of research within and across disciplines involved surveying academic journals in eleven disciplines to identify the types of research that are, typically, published in each. Six disciplines were then selected for inclusion in the corpus: philosophy, history, political science, applied linguistics, biology and physics. This selection was based on several competing factors, including the ability to:

- (i) Represent each article type with at least two disciplines;
- (ii) Represent disciplines with multiple article types to the extent that it was possible to do so;
- (iii) Include disciplines from a range of academic areas; and,
- (iv) Relate findings to previous disciplinary research.

	Theoretical	Qualitative	Quantitative	Total
Philosophy	280,826	–	–	280,826
History	–	282,898	–	282,898
Political Science	–	191,791	230,386	422,177
Applied Linguistics	–	237,089	202,871	439,960
Biology	–	–	154,824	154,824
Physics	194,029	–	183,279	377,308
Total	474,855	711,778	771,360	1,957,993

Table 1: Corpus description in number of words (thirty texts per discipline/category)

The second step in corpus compilation was to develop discipline-specific processes for sampling and accurately labelling articles according to the research paradigm reported on in the article. Thus, disciplinary experts were consulted to advise in the selection of high-quality journals to serve as the source for corpus texts, and to aid in the development and validation of discipline-specific operational definitions that would enable the accurate categorisation of articles by research type (see Gray, 2011).

Three types of primary research are included in the corpus: theoretical, qualitative and quantitative research. Theoretical articles have no observed data; rather, they present arguments to advance and explore theoretical concepts in the field. Qualitative research studies focus on the observation and description of empirical data, make little or no attempt to quantify that data, and are typically observational in nature. Quantitative articles (observational or experimental) also analyse observed data; however, that data is numerically based and often involves a comparison of groups. For the purposes of this paper, I will consider each research type a ‘register’ of academic writing.

Table 1 describes the final corpus. Each discipline is represented by only those research types that are prevalent in that discipline. Each discipline/register combination is represented by thirty research articles selected from eight to ten peer-reviewed journals that reflect a variety of areas or sub-disciplines within the field. The resulting 270-text corpus contains approximately two-million words, and enables comparisons of the same journal register across multiple disciplines (e.g., quantitative research in political science, applied linguistics, biology and physics), and, to a lesser extent, comparisons across two registers within a single discipline (e.g., qualitative and quantitative research in political science).

The corpus was ‘tagged’ using the Biber Tagger (see Biber, 1988; and Biber *et al.*, 1999) and evaluated for accuracy based on a subset of fifteen text samples. These samples were, first, manually coded for tagging errors. A computer program was written to correct systematic errors, and each sample

was then coded a second time to enable the calculation of rates for precision and recall. These rates, and a more detailed description of the tag-checking process, can be found in Gray (2011: Chapter 5 and Appendix B).

3. Conducting a new multi-dimensional analysis of cross-disciplinary research article registers

The normalised rates of occurrence (per 1,000 words) for more than 100 linguistic features were obtained using a tag-count program developed by Biber, which processes tagged texts and identifies major parts of speech, semantic classes of words, syntactic structures and stance features. Pilot factor analyses (using SPSS v. 19.0) identified meaningful combinations of features to be included in the final factor analysis. Seventy features with communalities of ≥ 0.250 and factor loadings of ≥ 0.30 on at least one factor in the pilot analyses were retained for the final factor analysis. (For a complete list of features with examples, see Gray, 2011: Chapter 9.)

Preliminary analyses confirmed that the final data set was appropriate for factor analysis (FA). The Kaiser–Olkin Measure of Sampling Adequacy ($KMO = 0.855$, meritorious) met the minimum requirement for FA. Bartlett's Test for Sphericity (approximate Chi-Square = 11005.23, $df = 2415$, $p = 0.000$) was significant. These tests indicate that adequate correlations exist in the data set, making FA an appropriate method for identifying correlation patterns in the data (Tabachnick and Fidell, 2007).

The final factor analysis was run using Principal Axis Factoring with Promax rotation (to allow some correlation between factors). The four-factor solution was determined to be the most interpretable when compared with three- and five-factor solutions. The four-factor solution resulted in Initial Eigenvalues (Total Variance Explained) that indicate a cumulative percentage of variance explained of 40.26 percent. All variables with a factor loading of 0.30 or above were considered important for that factor, and are summarised in Tables 4 to 8. Factor (or dimension) scores for each text were calculated using z -scores (see Biber, 1988: 94) to quantify the extent to which a text utilises the patterns of variation identified by each factor. Dimension scores for each sub-corpus consist of the mean dimension score of the thirty texts in that sub-corpus.

One-way ANOVAs show that all factors were significant ($\alpha < 0.05$). ANOVA results are presented along with r^2 values (see Conrad and Biber, 2001: 28) with the figures plotting the distribution of registers/disciplines along each dimension in Section 4. The Games–Howell procedure (equal variances not assumed) was used for post-hoc comparisons, although due to space constraints these results will not be presented here. In the next section, each dimension is described and interpreted in detail based on the composition of the dimension along with the distribution of disciplines and registers along the dimension.

<i>Positive features</i>	
Pronouns	nominal pronouns (0.69), pronoun <i>it</i> (0.62), first-person pronouns (0.58), demonstrative pronouns (0.52)
Nouns	nouns of cognition (0.57)
Adjectives	predicative adjectives (0.70), evaluative attributive adjectives (0.33)
Verbs	verb <i>be</i> (0.79), verb <i>have</i> (0.67), causative verbs (0.34)
Modal verbs	modals of prediction (0.69), modals of possibility (0.66), modals of necessity (0.65)
Adverbs	general adverbs (0.54), stance adverbs (0.47), adverbs of time (0.34)
Conjunctions	subordinating conjunction – conditional (0.83), adverbial conjuncts (0.48), subordinating conjunctions (0.39)
Finite clauses	<i>that</i> -clauses controlled by nouns of likelihood (0.65), <i>that</i> -clauses controlled by verbs of likelihood (0.59), <i>that</i> -clauses controlled by factive adjectives (0.48), <i>that</i> -clauses controlled by attitudinal nouns (0.47), <i>that</i> -clauses controlled by factive nouns (0.44), <i>wh</i> -clauses (0.34)
Non-finite clauses	<i>to</i> -clauses controlled by stance adjectives (0.37), <i>to</i> -clauses controlled by verbs of probability (0.32)
<i>Negative features</i>	
Nouns	nouns (−0.75), process nouns (−0.40)
Verbs	past tense verbs (−0.67)
Passives	passive postnominal modifiers (−0.53), agentless passive voice verbs (−0.32)
Other	prepositions (−0.39), type–token ratio (−0.35), word length (−0.31)

Table 2: Factor structure for Dimension 1: academic involvement and elaboration *versus* information density

4. Patterns of variation across discipline and research type

4.1 Dimension 1: academic involvement and elaboration *versus* information density

Dimension 1, labelled Academic Involvement and Elaboration versus Information Density, is composed of twenty-six positive features and eight negative features (Table 2). Positive features include various types of pronouns, modifiers (several types of adjectives and adverbs), explicit markers of logical and grammatical relationships (conditional and other subordinating conjunctions, and adverbial conjuncts), and structures that convey personal stance meanings (possibility/permission/ability modals, various *that*- and *to*-clauses controlled by stance nouns, verbs and adjectives), and cognition nouns. Features that loaded negatively on Dimension 1

included various noun-phrase structures (all nouns, process nouns, passive post-nominal modifiers and prepositions), agentless passive voice verbs, past tense verbs, type-token ratio and word length.

Despite the fact that this dimension has been extracted from a corpus of all written registers, there are a surprising number of similarities between this dimension and Biber's (1988) Dimension 1 (involved *versus* informational discourse) and Biber's (2006) Dimension 1 (oral *versus* literate texts), both of which consistently revealed variation across spoken and written registers.

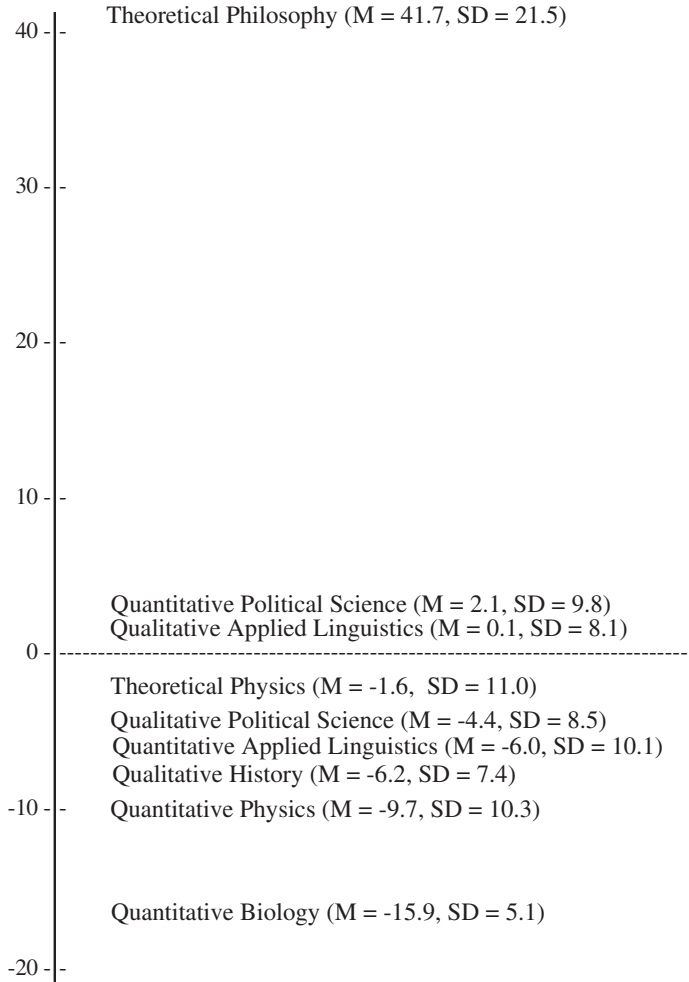
On the one hand, the negative Dimension 1 features in this study are closely aligned with Biber's (1988, 2006) Dimension 1 negative features, which he has associated with an informational purpose. The text examples (discussed below) illustrate that these features can also be associated with highly compressed styles of discourse in which a great deal of information is packed into dense noun-phrase structures (e.g., see Biber and Gray, 2010, 2013; and Biber *et al.*, 1999).

The positive features on Dimension 1, on the other hand, differ in important ways from Biber's earlier dimensions. Many pronouns, stance markers and clausal features load positively on the present Dimension 1. Biber (1988, 2006) has associated similar types of features with elaborated and 'involved' language production. However, many of Biber's 1988 and 2006 positive Dimension 1 features associated with interactional spoken language were eliminated during the preliminary FAs because they did not exhibit variation in the RA corpus (e.g., contractions, indefinite pronouns, *wh*-questions, *that*-deletion and discourse particles; see Biber, 1988, 2006; and Friginal, 2009). Thus, apart from these conversational features, Dimension 1 reflects a distinction that has been consistently identified in MD analyses: that of high-density informational language and more involved language production. However, Dimension 1 also offers a more precise picture of involvement—one that is specifically focussed on features that play an important role in creating elaboration and involvement in academic writing.

Figure 1 displays the distribution of the disciplines and article types along Dimension 1, and reflects the dichotomy between theoretical philosophy and all other disciplines/registers. Philosophy has an extremely high positive dimension score, while all other disciplines range from near-zero to highly negative scores. The quantitative hard sciences (biology and physics) exhibit the lowest scores, while the political science, applied linguistics and history registers exhibit a more balanced use of the positive and negative features.

The positive features on Dimension 1 serve several functions related to involvement. First-person pronouns in particular are used to bring the author, the reader and often humankind in general (as in Example 1) into the discourse, creating a sense of interaction and relating points and ideas to the readers as members of humankind. Other features, such as *it* and many of the clausal features, are used to convey stance (particularly in

ACADEMIC INVOLVEMENT & ELABORATION



INFORMATIONAL DENSITY

Figure 1: Distribution of factor scores along Dimension 1: academic involvement and elaboration *versus* informational density. One-way ANOVA: $F = 66.62$, $p = 0.000$, $r^2 = 0.67$

Example 2). These overt evaluations, along with other types of subordination, explicitly mark connections, meaning relationships and interpretations, and form the basis for the author's argument. The prevalent use of cognition nouns may reflect the nature of philosophy itself, as a discipline that relies upon arguments and logic (both mental processes) to explore the human experience.

Examples 1 and 2 illustrate a heavy reliance on the use of positive Dimension 1 features in philosophy articles (see Table 2; single-word features such as first-person pronouns, cognition nouns and adverbs are shown in boldface; verb features are underlined, and clausal features are enclosed in square brackets with controlling words/subordinators double underlined).

Positive Dimension 1 features in theoretical philosophy

- (1) Unmotivated **desires**, by contrast, ARE states or dispositions which lack this basis in **reasons**. They ARE states with which **we simply find ourselves**. . . The **concept** of an unmotivated state or disposition seems essential₁ [to the ordinary prephilosophical **notion of desire**]₁. Not only IS desire not something we reason our way to, but once₂ [**desire exists**]₂ it IS generally not sensitive to reasons in the way other **attitudes ARE**.
- (2) Well, **it MAY** prove a problem for Wollheim's claim₂ [that **it IS impossible**₃ [to isolate **memories** from their psychological context,]₃]₂ but **it** does not undermine our weaker claim₄ [that **memories ARE** weakened when they ARE isolated in this way]]₄. Martin does not tell us [how much of the Spanish conversations he **now** remembers,] but **it** seems plausible₅ [to suppose₆ [that the **fact**₇ [that he no longer understands Spanish]₇ **will HAVE** resulted in his **now** not being able₇ [to recall some of the details of the conversations.]₅]₆]₇

In contrast, quantitative physics and quantitative biology texts have highly negative dimension scores, exhibiting a relative lack of positive Dimension 1 features and a dense use of negative Dimension 1 features. The following excerpt illustrates how the negative features result in informationally dense discourse, in which long noun phrases are created by prepositional phrase and passive post-nominal modifiers, and nouns as noun pre-modifiers. In Example 3a, negative Dimension 1 features are marked as follows: all nouns are shown in boldface, prepositions are underlined, and passive structures are double underlined.

Negative Dimension 1 features in quantitative physics

- (3a) The **signal count** produced by the laser is then proportional to k, with j as photon density and k as the number of photons used to produce the signal in the fragmentation process. Due to the large **number of results described here**, we make a stepwise and systematic **approach**, proving each **step** carefully even if they might have been proved previously in other **publications** (always referenced).

When positive Dimension 1 features occur near the end of the passage, they allow the authors to make connections outside the actual events of the

research (see Example 3b, in which the excerpt is reproduced with positive features highlighted). This is in contrast to the use of positive features seen in Examples 1 and 2, above, in which the positive Dimension 1 features were used throughout the discourse to comment on the research as well as to form the argument and present evidence.

Positive Dimension 1 features in quantitative physics
(marked as per Examples 1 and 2, above)

- (3b) The signal count produced by the laser **IS then proportional** to k , with j as photon density and k as the number of photons used to produce the signal in the fragmentation process. Due to the large number of results described here, **we** make a stepwise and systematic approach, proving each step **carefully** even **if**₁ [they MIGHT have been proved **previously** in other publications (**always** referenced).]₁

In contrast, the following excerpt, from an article on quantitative biology, shows an almost complete lack of positive Dimension 1 features. Although the actions described in Example 4 had human agents, passive structures are favoured over first-person pronouns.

Negative Dimension 1 features in quantitative biology methods
(marked as per 3b, with past tense italicised)

- (4) Stable **isotopes of carbon and nitrogen** were used to estimate **food-chain length** and identify **patterns of material flow** through dominant trophic **pathways** for each food web (Hoeinghaus et al. 2007a). The aquatic **food webs** ANALYZED in this study were delimited by fishes as consumers plus their aquatic and riparian **prey** and **production sources** CONSUMED throughout the web leading to those consumers.

Of course, philosophy, biology and quantitative physics illustrate the two extremes of Dimension 1. Mean dimension scores for the remaining disciplines and registers cluster within a six-point range near zero (see Figure 1) and show that they rely on elaborating features and information density features in more balanced ways than philosophy, biology and quantitative physics. The following excerpt from qualitative applied linguistics illustrates this balance (positive features are shown in boldface, and negative features are underlined). The example exemplifies both the frequent use of nouns, prepositions and passives (which maintain the nominal style of academic writing and illustrate negative features) alongside the use of positive features such as cognition nouns (*theories* and *concepts*), BE verbs, predicative adjectives, modal verbs, adverbs, adverbial conjuncts (*however*) and stance *to*-clauses ('crucial to identify X').

Positive and negative Dimension 1 features in qualitative applied linguistics

- (5) **Theories** of social actions and learning/socialization such as Lev Vygotsky's (1978) **are useful in helping teachers** create practical situations in which writing students **can learn in** social situations. To understand **how this learning happens, it is crucial**₁ [to identify the interactional details of group work/discourse]₁. Most of the sources that investigate writing students' talk (some of which **are covered** below), **however**, tend to focus on **purely** theoretical **concepts**, the social power structures inherent in tutor/peer relationships, or a priori analyst-imposed categories of group work talk.

The dimension scores for two registers are worthy of special mention, as they counter the initial perception that Dimension 1 distinguishes between the natural sciences and the non-science disciplines. Theoretical physics actually has a dimension score (−1.6) that is more similar to the political science and applied linguistics registers (which range from 2.1 to −6.0) than the other hard science registers/disciplines. In theoretical physics, arguments are based on a logical progression of evidence. Examples 6 and 7 illustrate the way that many positive Dimension 1 features (shown in boldface) are employed in order to engage and guide the reader through the progression of the analysis. First-person pronouns, time adverbs and demonstrative pronouns function to keep the reader involved and focussed on the relevant portion of the analysis, while conditional subordinators (e.g., *if*), adverbial conjuncts and modal verbs specify logical relationships and the conditions upon which the analysis is based. At the same time, negative Dimension 1 features (which are underlined, except for nouns) are still prevalent and maintain a relatively impersonal and technical focus.

Positive and negative Dimension 1 features in theoretical physics

- (6) In particular, **we** give the constraints which the function {formula} **should** satisfy and given this function **we** define functions {formula}, for all even characteristics {symbol}, which satisfy the constraints from Section 2.3. Let {symbol} **be** the subgroup of {symbol} which fixes the characteristic {formula}:
 {formula}
For {symbol} **we required** that {formula}, that **is**, {formula} a modular form **on** {symbol} **of** weight 8. Given such a modular form {formula} **we now** define, **for** each even characteristic a function {formula} **in** such a way that Eq. (3) holds.

Positive and negative Dimension 1 features in theoretical physics

- (7) **It is difficult**₁ [to accommodate the result of NA with those of the other experiments]₁. **It is also hard**₂ [to see [₃ **how** the

outcome of NA could be reconciled with the results obtained in the dispersive analysis]₃]₂. Furthermore, if the numbers given by ISTR+ (obtained from K + decays) and NA (extracted from KL decays) were both **true**, **this would** signal an enormous isospin violation in the scalar form factors of K.

It is also somewhat surprising that qualitative history has a lower dimension score (−6.2) than any of the non-science disciplines and registers (see Figure 1). Example 8 illustrates how qualitative history reports rely extensively on some of the negative Dimension 1 features, including dense noun-phrase structures (often with prepositional phrases functioning as noun modifiers) and the use of past tense to describe and report events. In contrast, many of the modal verbs and stance-conveying structures represented by positive Dimension 1 features, and the passive structures that are common in natural science texts, are absent from these narratives as the author presents a detailed, yet relatively objective, account of past events. (Negative Dimension 1 features are marked as follows: nouns are shown in boldface, prepositions are underlined and past tense verbs are italicised.)

Negative Dimension 1 features in qualitative history

- (8) The **Indian Inquisition** *ended* when the **Council** of the **Indies** *revoked* the **bishop's** inquisitional **powers** in 1543. In 1571, when **Philip** II formally *established* a Holy **Office** in **New Spain**, he specifically *prohibited* **trials** against indigenous **colonists**. Most **historians** have attributed this **decision** to the **failure** of the earlier **Indian Inquisition**.

To summarise, at the most basic level, Dimension 1 illustrates a divide between philosophy and other disciplines/register. In addition, Dimension 1 provides some evidence that quantitative natural science research relies on features that are related to informational density to a greater extent than social science and humanities research, and this finding accords with recent findings by Biber and Gray (2013) that the dense nominal style of academic writing has been adopted to a much greater extent in science writing than in non-science writing over the past 100 years. However, another noteworthy contribution of Dimension 1 to explaining disciplinary and register variation is that individual disciplines/register (e.g., history and theoretical physics) exhibit distinctive patterns of use based on unique characteristics, such as the nature of evidence, the process by which arguments are proved (e.g., through logical progressions or narrative descriptions) and the topic of the research.

4.2 Dimension 2: contextualised narration *versus* procedural discourse

Twenty-two positive features and six negative features make up the second dimension, labelled Contextualised Narration *versus* Procedural Discourse

<i>Positive features</i>	
Pronouns	third-person pronouns (0.65)
Nouns	group nouns (0.49), nominalisations (0.32), animate nouns (0.43)
Adjectives	topical attributive adjectives (0.53), attributive time adjectives (0.47)
Verbs	past tense verbs (0.55), aspectual verbs (0.52), perfect aspect verbs (0.48), communication verbs (0.47), present progressive verbs (0.42)
Conjunctions	phrasal co-ordinating conjunctions (0.51), clausal co-ordinating conjunctions (0.35)
Finite clauses	<i>that</i> -relative clauses (0.46), <i>that</i> -clauses controlled by non-factive verbs (0.45), <i>wh</i> -questions (0.32)
Non-finite clauses	<i>to</i> -clauses controlled by verbs of modality, causation and effort (0.57), <i>to</i> -clauses controlled by verbs of desire (0.41), <i>to</i> -clauses controlled by stance nouns (0.35)
Other	word length (0.52), word count (0.36), type–token ratio (0.31)
<i>Negative features</i>	
Nouns	technical nouns (−0.61), quantity nouns (−0.46), concrete nouns (−0.37)
Adjectives	attributive adjectives indicating size (−0.37)
Passives	agentless passive voice verbs (−0.52), passive voice verbs with <i>by</i> -phrases (−0.47)

Table 3: Factor structure for Dimension 2: contextualised narration *versus* procedural description

(see Table 3). Positive features include references to human beings (third-person pronouns, group nouns and animate nouns), indicators of time and topic (time and topic adjectives, and various tense and aspect markers), structures indicating communicative acts (communication verbs, *that*-clauses headed by communication verbs), *that*-relative clauses, and phrasal and co-ordinating conjunctions. Negative features on Dimension 2 include three types of nouns (technical, quantity and concrete nouns), size adjectives and passive voice verbs (agentless and *by*-passives).

Figure 2 shows that research type (and to a lesser extent discipline) plays an important role in how the disciplines and registers are distributed along Dimension 2 (in contrast to Dimension 1, where the two poles of the dimension more closely reflected the nature of the discipline). That is, Dimension 2 can be associated with the primary way in which evidence is presented and discussed.

The three qualitative registers (history, political science and applied linguistics) group together on the high positive end of the dimension, and all rely on the narration of events as a form of evidence (e.g., a description of participants' actions, thoughts or reflections during study events in qualitative applied linguistics, or an account of events and actions during a

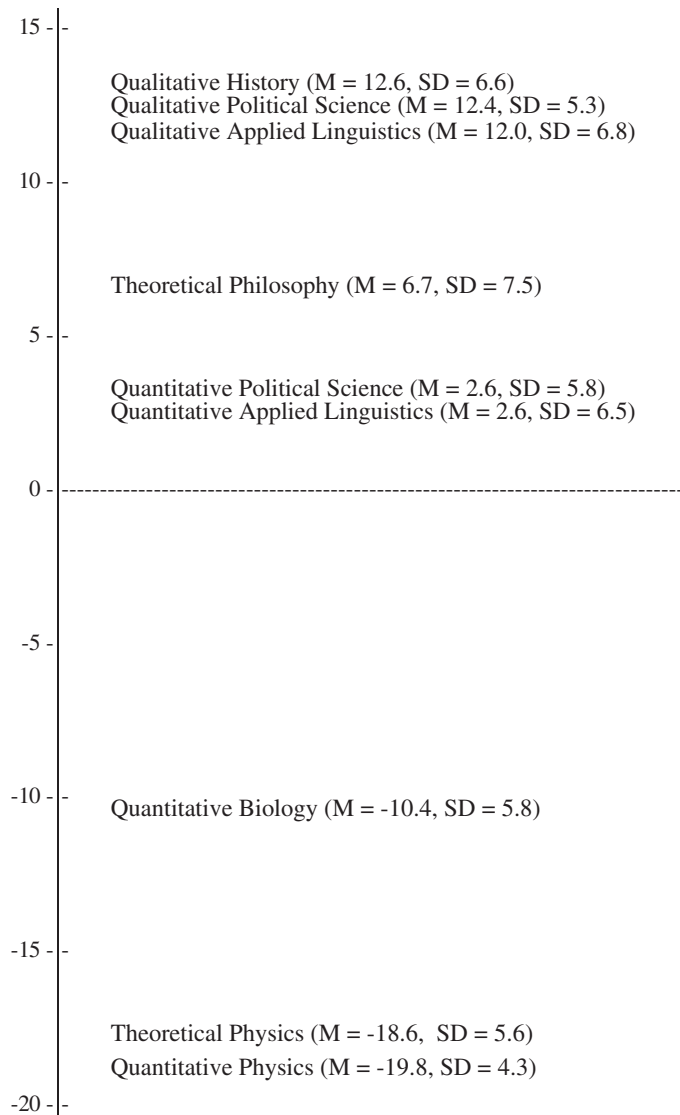
CONTEXTUALIZED NARRATIVE DESCRIPTION**PROCEDURAL DESCRIPTION**

Figure 2: Distribution of factor scores along Dimension 2: contextualised narrative description *versus* procedural description. One-way ANOVA: $F = 138.99$, $p = 0.000$, $r^2 = 0.81$

particular time period or political situation in history and qualitative political science).

The positive features for Dimension 2 help create this narrative discourse, particularly third-person pronouns, human/animate nouns, past tense and perfect aspect verbs, and communication verbs (see, for example, Biber, 1988: Dimension 2; and Biber, 2006: Dimension 3). In Examples 9 and 10, the focus is on reconstructing an event to serve as the foundation for interpretations and subsequent claims. Past tense, perfect and progressive aspect, aspectual verbs, and time adjectives help to establish the sequence and timing of events and happenings (*italicised*), while communication verbs (*underlined*) are used to report speech from the often human actors represented by third-person pronouns, group nouns and animate nouns (*boldface*). Nominalisations (*underlined*) function to turn actions and processes into nominal groups, allowing for the representation of past action in noun phrases. Taken together, the positive Dimension 2 features allow qualitative researchers to reconstruct series of events (Example 9), which they then interpret and draw conclusions from (Example 10).

Positive Dimension 2 features in qualitative political science (Example 9) and qualitative applied linguistics (Example 10)

- (9) The Charkaoui v. Canada case involves two permanent **residents** and one **refugee**, all of Arab origin. All *WERE* detained indefinitely pending a deportation₁ [that may never happen,]₁ on the authority of ministerial “security certificates” based on secret evidence₂ [that *LED* **government** to believe that **they** might be dangerous.]₂ None of **them** *WERE* even suspected of *HAVING COMMITTED* a crime in Canada or elsewhere.
- (10) A possible interpretation of this note is that **he** *REALIZED* that **he** *HAD MANAGED* to overcome this error in the Month 4 presentation. This is interesting as it *SUGGESTS*₁ [that **he** *HAD BEEN* conscious even of his target-like use of garbage]₁. ... Katsu, consciously reflecting on his target-like performance, *SEEMED* to demonstrate a more positive form of noticing; that is, that **he** *HAD GAINED* new knowledge, filling a previous gap in his L. Katsu’s note therefore gives a further **indication**₂ [that the countability of the form *HAS BEEN FURTHER INTEGRATED* into his interlanguage system.]₂

The negative end of Dimension 2 (characterised by technical, quantity and concrete nouns, as well as passive voice) indicates a focus on procedural activities – describing research methods and reporting the outcomes of those procedures. Quantitative biology and both physics registers have negative Dimension 2 scores. In theoretical physics (see Example 11), evidence is often presented as a series mathematical formulae (procedures), and technical and quantity nouns are used to refer to related concepts (*boldface*). Passives

(underlined) are also used to introduce and explain formulas, as well as to establish procedural steps related to the computations.

Negative Dimension 2 features in theoretical physics

- (11) Therefore, S_z is a conserved **quantity**. As a result, in terms of the standard basis vectors $\{\text{symbols}\}$ and $\{\text{symbol}\}$, the reduced two-spin density matrix, which is constructed by tracing out the rest spin degrees of freedom, can be written as $\{\text{formula}\}$. In terms of the spin correlation functions, the elements of $i, i+1$ in **equation (2)** are given by $\{\text{formula}\}$.

In quantitative physics and biology, technical, quantity and concrete nouns are used to discuss specific methods and results, as objects under study are described in these terms and then summarised in quantitative displays, such as figures and tables. The passive is used to describe methodological steps and processes, indicate the results of procedures, and refer to quantitative data displays.

Negative Dimension 2 features in quantitative physics (Example 12) and quantitative biology (Example 13)

- (12) Four **data samples** each are selected from the core and segment **data** sets. The **data samples** are defined by the energy measured in the core and are labeled: DEP: The **sample** contains events with a core **energy** in the region of (1593 ± 5) keV.
- (13) The interpretation of NMDS and PCA axes used in this analysis is presented in Table 1, and details of their loadings can be found in Appendices B and C

Both quantitative political science and applied linguistics exhibit low positive dimension scores on Dimension 2, indicating a reliance on both positive and negative features. Two competing factors may contribute to this finding. First, the quantitative research methodology that these studies employ make the negative features discussed above also important for describing study methods and reporting quantitative results, as in Example 14. Thus, regardless of discipline, quantitative research utilises concise, informative descriptions of the procedures used to collect and analyse data, while maintaining a focus on those procedures (rather than the researcher).

Negative Dimension 2 features in quantitative applied linguistics

- (14) The 31 frequent and 31 infrequent collocations were combined, in random order, and attached to the collocation instrument. Participants were asked to rate all 62 collocations on the basis of their commonness in the English language...

However, political science and applied linguistics also share an underlying focus on human subjects or events (which inherently involve people and their actions). They also tend to provide extensive introductions to place themselves in an existing body of research (often in the form of a narration of the situation and previous research). Example 15 illustrates this more narrative-type research introduction (positive Dimension 2 features marked as per Example 9, with relative clauses also marked).

Positive Dimension 2 features in quantitative political science

- (15) The transformations of the composition of the workforce, of the **student population**, and of the electorate are among the most significant changes₁ [that *HAVE ACCOMPANIED* the population change of the past 20 years.]₁ It is particularly important to recognize the role₂ [that Latinos and specifically Latino **immigrants** play in these transformations.]₂ Both **scholars** and policy makers recognize the significant role₃ [that Latinos *PLAYED* in the transformation of the workforce and the education]₃ and *HAVE CONSIDERED* the consequences of this change (REFS).

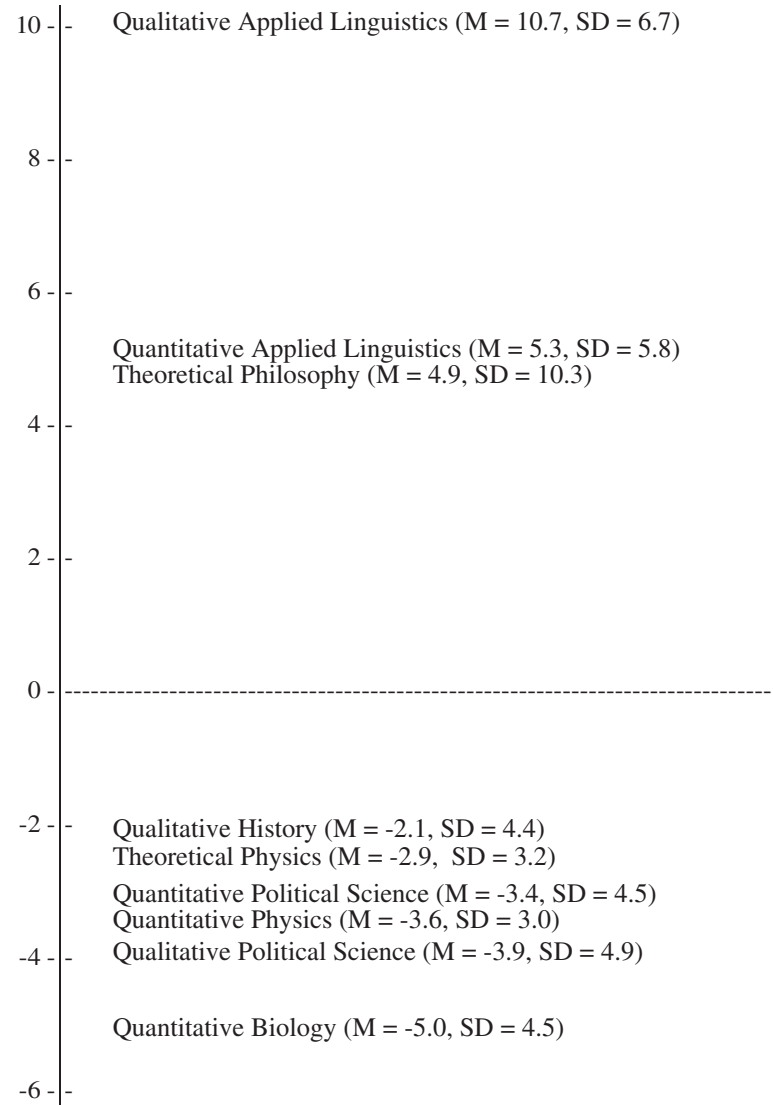
4.3 Dimension 3: human *versus* non-human focus

Interpreted to represent a human versus a non-human focus, Dimension 3 is characterised by eleven positive features and four negative features (Table 4). Positive features include second- and third-person pronouns, process nouns, three types of verbs that typically have human agents (mental, activity and communication), and progressive aspect. Clausal features such as *wh*-clauses and many stance-conveying clauses are also included in the positive features. The four negative features include attributive adjectives, topic adjectives, adverbs and prepositions.

Figure 3 again shows a dichotomy in the distribution of the register/discipline dimension scores, this time between disciplines whose subject matter is that of human beings and their mental and cognitive activities (both applied linguistics registers and theoretical philosophy) and all other disciplines. In applied linguistics, researchers are often concerned with the ways in which language is used, how language is learnt and acquired, and the effectiveness of teaching methods to promote language acquisition and development. While philosophy is not specifically focussed on language, it is focussed on the human mind and how we use logic and morals to understand human nature, human problems and human cognition (to name a few).

The following excerpts illustrate how the positive features on Dimension 3 correspond to this human focus: second- and third-person pronouns to refer to people (italicised) and several verbal structures that take human agents (mental, activity and communication verbs are

Human Focus



Non-human Focus

Figure 3: Distribution of factor scores along Dimension 3: human *versus* non-human focus. One-way ANOVA: $F = 28.42$, $p = 0.000$, $r^2 = 0.47$

<i>Positive features</i>	
Pronouns	second-person pronouns (0.40), third-person pronouns (0.35)
Nouns	process nouns (0.50)
Verbs	mental verbs (0.65), activity verbs (0.60), communication verbs (0.51), present progressive verbs (0.49)
Finite clauses	<i>that</i> -clauses controlled by factive verbs (0.42), <i>wh</i> -clauses (0.33)
Non-finite clauses	<i>to</i> -clauses controlled by verbs of desire (0.46), <i>to</i> -clauses controlled by speech verbs (0.45)
<i>Negative features</i>	
Adjectives	attributive adjectives (−0.54), attributive adjectives indicating topic (−0.42)
Adverbs	general adverbs (−0.31)
Other	prepositions (−0.35)

Table 4: Factor structure for Dimension 3: human *versus* non-human focus

shown in boldface, and stance clauses controlled by two uniquely human processes—communication and desire—are shown in square brackets]. Within the verbal structures, the human agents can be explicitly used as the subject (verbs marked 2, 3, 4) or implied based on the discourse context (verbs marked 1, 5).

Positive Dimension 3 features in applied linguistics (Example 16) and theoretical philosophy (Example 17)

- (16) Teachers built background knowledge by **asking**_{1 – implied ‘teachers’} learners [what *they* **thought**_{2 – explicit ‘learners’} the site would be about]
- (17) Egalitarians can be pluralists about value. *They* **think**_{3 – explicit egalitarians as ‘they’} that there is a reason to level down—where doing so will make a *distribution* fairer—but **recognize**_{4 – explicit egalitarians as ‘they’} that there are also reasons not to... The (non-person-affecting) reason to level down is unlikely to outweigh the reasons there are to **prefer**_{5 – implied ‘we’ as humankind} people [to be better rather than worse off.]

In quantitative and qualitative applied linguistics, mental, activity and communication verbs, and stance clauses controlled by desire and speech verbs, are often used to report the thoughts, speech and actions of participants in the research (Examples 17 and 18), while in theoretical philosophy, these same verbs are used to talk about humankind more generally (Example 19) or unreal characters (Example 20) who are used to explore and illustrate logical processes.

Positive Dimension 3 features in applied linguistics (Examples 17 and 18) and philosophy (Examples 19 and 20)

- (17) Regarding the kind of difficulties *they* **experienced**, some **commented** on the inferential reading questions and writing conventions. Many L students **mentioned** vocabulary as one of the key difficulties in **taking** the test. [Qualitative]
- (18) After **reading** these instructions, each participant completed a series of practice sentences (see Appendix D) and **made** practice grammaticality judgements for these sentences. [Quantitative]
- (19) Memory is ordinarily taken to be factive. One cannot **remember** that which did not happen.
- (20) Mary **knows** all there is to **know** about physics, chemistry and neurophysiology, yet has never **experienced** colour. Most philosophers **think** that if Mary **learns** something genuinely new upon seeing colour for the first time, then physicalism is false.

On first consideration, it may seem surprising that qualitative history and both political science registers have negative rather than positive Dimension 3 scores, as the study of historical events and social organisation are inherently connected with human beings and their actions. However, the following examples show that the objects of study in history and political science are typically events and situations that make up human history and society, rather than specific human beings and their cognitive processes or communicative activities.

Focus on non-human entities in history (Example 21) and political science (Example 22)

- (21) Women's battles and breakthroughs in the liberal professions during the Third Republic—their struggle for entry in the late nineteenth century, their growth and progress, and the resistance they encountered anew in the 1930s—are one chapter in a bigger story of professional exclusion.
- (22) Critically, decommodification reflects the quality as well as quantity of social rights and entitlements; the mere presence of social assistance or insurance may not necessarily bring about significant decommodification if they do not substantially emancipate citizens from market dependence.

To summarise, Dimension 3 distinguishes between disciplines with human beings and their thoughts, knowledge and communicative practices at the heart of the discipline's subject matter, and disciplines which focus on either historical events and states (history and political science) or the non-human natural world (biology and physics).

<i>Positive features</i>	
Nouns	nominalisations (0.43), process nouns (0.38), other abstract nouns (0.32)
Adjectives	relational attributive adjectives (0.45)
Verbs	existence verbs (0.37)
Finite clauses	<i>that</i> -clauses controlled by likelihood adjectives (0.38), <i>to</i> -clauses controlled by stance adjectives (0.33)
Other	word length (0.58)
<i>Negative features</i>	
Adverbs	time adverbials (−0.47)

Table 5: Factor structure for Dimension 4: ‘academese’

4.4 Dimension 4: ‘academese’

Eight positive features and one negative feature load on Dimension 4 (see Table 5). Positive features include three types of abstract nouns (process nouns, other abstract nouns and nominalisations), existence verbs and several adjective-based structures. Given that relatively few features load on this dimension, interpretations are more limited than for the first three dimensions.

Figure 4 shows that only two disciplines have positive Dimension 4 scores: applied linguistics and political science. In both disciplines, quantitative research reports demonstrate the highest positive score, followed by qualitative research reports. The label ‘academese’ has been assigned to the positive end of this dimension, based equally on this distribution, an analysis of the non-linguistic characteristics of the disciplines, and the linguistic composition of the dimension. Taking all three factors into account, a positive Dimension 4 score appears to reflect a concern to overtly represent research as empirical, well-motivated and founded in previous research. The negative end of Dimension 4 is un-interpretable with only one feature.

Comparing the non-linguistic characteristics of the only two disciplines with positive Dimension 4 scores with all other disciplines shows that applied linguistics and political science were the two disciplines that most consistently included integrated citations, clearly labelled data and research methods, and explicitly stated research questions and design (Gray, 2011: Chapter 6). Thus, these two disciplines are careful to express overtly academic (and scientific) values (e.g., grounding in prior theory and research, and research design) through very explicit means.

This was in contrast to qualitative history and quantitative physics, which are also both empirical (as conclusions and interpretations are drawn based on observed data). However, academics in these two fields often do not overtly refer to the field and its research practices as ‘empirical’.² On

² T. Porter, personal communication, 6 May 2009; and G. Lubick, personal communication, 26 January 2010.

'Academese'

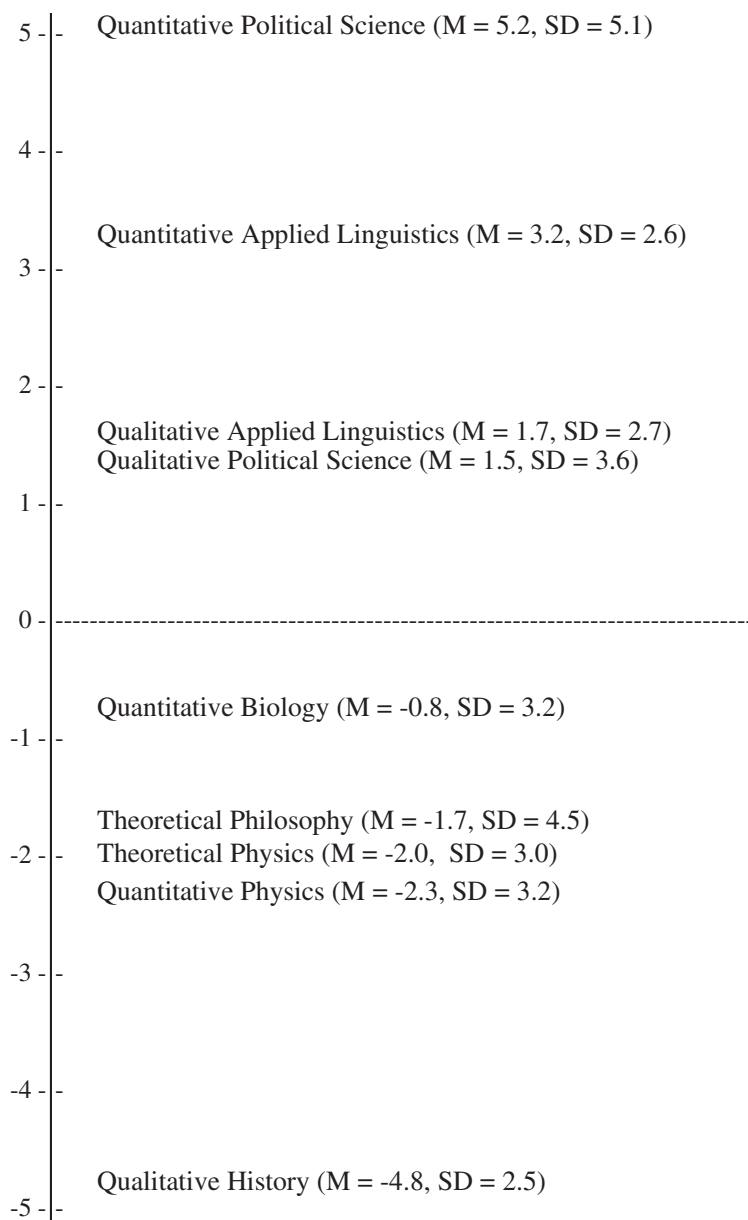


Figure 4: Distribution of factor scores along Dimension 3: 'Academese'. One-way ANOVA: $F = 24.33$, $p = 0.000$, $r^2 = 0.43$

the other hand, the theoretical registers (philosophy and physics) are not empirical in nature. In contrast, biology research is widely understood to be empirical, and spends less time making elaborate connections to previous research (evidenced in part by the shorter nature of biology texts as well as the non-linguistic analysis of the sub-corpus). Rather, many of these values are understood implicitly rather than being reflected explicitly in the language of the texts.

In addition, the linguistic composition of this dimension also supports an interpretation connecting it to the practice of explicitly packaging the research activities of the field as ‘scientific’ and ‘empirical’ inquiry based on a strong foundation. The extensive use of abstract and process nouns, along with nominalisations, reflects Halliday’s notion of grammatical metaphor, which ‘allows any observation, or series of observations, to be restated in summary form’ (Halliday, 2004: 20) in nominal structures. Halliday argues that such grammatical metaphor is important in scientific language in order to position observations as entities that can be used as the foundation for the next steps in the research.

As applied linguistics and political science extensively build on existing theories and findings, they rely on many of these structures, representing concepts and processes in nominal forms as they place their research motivations and interpret their findings (illustrated in Example 23, in which nominalisations, abstract nouns and process nouns are shown in boldface). In particular, quantitative registers within these two disciplines utilise these structures to an even greater degree, as they also seek to reflect the scientific foundations of their methods.

Nominalisations, abstract nouns and process nouns in quantitative political science

- (23) The underlying logic of segmented **mobilization** is that the contextual **factors** that significantly altered the **patterns** of Latino **participation** since the mid-1990s also had an **impact** on the **receptivity** of segments of the Latino **electorate** to **mobilization** requests.

Other positive Dimension 4 features are used extensively to describe and interpret the results of the research, illustrated in Examples 24 and 25. For example, relational adjectives (underlined) allow for comparisons of results, while existence verbs (double underlined, for example, *appear*, *contain*, *defined*, *exist*, *illustrate*, *include*, *indicate*, *reflect*, *represent*, *stay*, *tend* and *vary*) introduce relationships and interpretations. Two types of stance clauses (*to*-clauses headed by stance adjectives, and *that*-clauses controlled by adjectives expressing likelihood) allow the researcher to hedge their interpretations and offer evaluations of their findings.

Positive Dimension 1 features in quantitative applied linguistics

- (24) This may INDICATE a **relationship** between modified output and noticing, similar to other recent empirical **findings** pointing to a **relationship** between noticing and learning in interactional **contexts** (REFS) ... **Claims** have also been made that recasts and **negotiation** may benefit L **development** in different ways, for example through positive **evidence** or corrective feedback or different response types (REFS).
- (25) All round then, given the various terms and working **definitions** used in **studies** of recasts, it may be **unhelpful**_{1,2} [to lump together all the different types under the single label of 'recasts']₁ and also [to assume that all recasts are categorically implicit.]₂

Although the interpretation of Dimension 4 is preliminary, this dimension appears to distinguish between disciplines which are carefully packaged as empirically based and employing empirical methods, rather than those which are not empirical in nature (theoretical articles in physics and philosophy), those which are understood to be implicitly empirical (quantitative physics and biology), and those which are not concerned with being overtly labelled as empirical (qualitative history). The linguistic features on this dimension thus help researchers in applied linguistics and political science to characterise and make sense of study motivations and results in order to offer interpretations of that data.

5. Concluding remarks: more than discipline

The new MD analysis reported in this paper has revealed that discipline is only one of the important characteristics of a research article that corresponds to linguistic variation. Rather, variation occurs along multiple parameters that have generally been unrecognised in previous studies of disciplinary variation. These parameters have included the nature of evidence, the presence or absence of data, qualitative *versus* quantitative *versus* theoretical research paradigms, and the object of study. Each of these characteristics corresponds to specific linguistic resources, which function together to create meaningful discourse that reflects the nature of the discipline and the research paradigm of the scholarly work.

In addition, the very different distributions along the four dimensions exhibited by the disciplines and RA registers highlights the complex nature of academic writing. Writers utilise a multitude of linguistic resources that correspond to an equally complex multitude of situational characteristics. For example, articles on theoretical physics simultaneously follow linguistic traditions of the natural sciences (procedural discourse and a non-human focus in Dimensions 2 and 3), while also creating involved forms of

argument in which the reader is engaged with the procedural steps that are presented as evidence of the author's findings (Dimension 1). Likewise, quantitative applied linguistics texts simultaneously exhibit the moderately dense informational focus of scientific writing (Dimension 1), a certain degree of narrative orientation (Dimension 2) and a heavy focus on human actors (Dimension 3).

Somewhat surprisingly, it appears that one of the dimensions of variation (Dimension 1) continues to highlight the differing degree to which texts balance the processes of elaboration/involvement and informational density. This finding comes despite the much narrower domain of inquiry into sub-registers that all belong to a larger, informational written register. While the specific features that loaded on this feature varied slightly from previous studies, these features reflect the specific involved *versus* informational functions that are most important for professional academic writing and add further support to the universality of this dimension of variation.

While previous research has largely disregarded the influence of the nature of research in studies of linguistic variation in research articles, this study has highlighted that such distinctions do matter. These findings are of particular importance to corpus-based studies on the characteristics of disciplinary writing. While it is probably not feasible (or perhaps necessary for the research goals) for every study of the language of a particular discipline to make categorical distinctions between sub-types of research articles, the results of this study support the claim that at least some attention to this matter is warranted in order to minimise the effect of register differences on findings. For example, the corpus design can be limited to one research type that is prevalent in a particular discipline, allowing for a specific sub-register to be focussed on. A second option would be to take care that a sample of research articles from any given discipline includes a variety of research types from the discipline, insuring representation of those research paradigms (without requiring a perfectly balanced representation). This would allow for a description of the discipline as a whole, and would also acknowledge and account for intra-disciplinary variation in creating meaningful analyses of language use in the disciplines.

This study has shown that an awareness of many different situational characteristics of research articles can augment our understanding of academic language within and across disciplines. Not only can this awareness help corpus linguists build more representative corpora, but it can also help us interpret patterns of linguistic variation and enable richer descriptions of language use. That is, each discipline and research paradigm makes use of linguistic resources in very specific ways. These patterns can be uncovered by carrying out systematic comparisons that reveal the underlying characteristics (such as topic, nature of evidence, object of study, methodological orientation and disciplinary practices) that correspond to specific patterns of language use.

References

- Afros, E. and C.F. Schryer. 2009. 'Promotional (meta)discourse in research articles in language and literary studies', *English for Specific Purposes* 28 (1), pp. 58–68.
- Banks, D. 2005. 'On the historical origins of nominalized process in scientific text', *English for Specific Purposes* 24 (3), pp. 347–57.
- Bazerman, C. 1994. *Constructing Experience*. Carbondale, Illinois: Southern Illinois University Press.
- Belcher, D. and A. Hirvela. 2005. 'Writing the qualitative dissertation: what motivates and sustains commitment to a fuzzy genre?', *Journal of English for Academic Purposes* 4 (3), pp. 187–205.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, D. and B. Gray. 2010. 'Challenging stereotypes about academic writing: complexity, elaboration, explicitness', *Journal of English for Academic Purposes* 9 (1), pp. 2–20.
- Biber, D. and B. Gray. 2013. 'Being specific about historical change: the influence of sub-register', *Journal of English Linguistics* 41 (2), pp. 104–34.
- Biber, D., B. Gray and K. Poonpon. 2011. 'Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?', *TESOL Quarterly* 45 (1), pp. 5–35.
- Biber, D., S. Conrad, R. Reppen, P. Byrd and M. Helt. 2002. 'Speaking and writing in the university: a multidimensional comparison', *TESOL Quarterly* 36 (1), pp. 9–48.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Conrad, S. and D. Biber. 2001. *Variation in English: Multi-dimensional Studies*. London: Longman.
- Diani, G. 2008. 'Emphasizers in spoken and written academic discourse: the case of *really*', *International Journal of Corpus Linguistics* 13 (3), pp. 296–321.
- Fang, Z., M. Schleppergrell and B. Cox. 2006. 'Understanding the language demands of schooling: nouns in academic registers', *Journal of Literacy Research* 38 (3), pp. 247–73.
- Friginal, E. 2009. *The Language of Outsourced Call Centers: A Corpus-based Study of Cross-cultural Interaction*. Amsterdam: John Benjamins.

- Gray, B. 2011. Exploring Academic Writing through Corpus Linguistics: When Discipline Tells Only Part of the Story. Unpublished PhD thesis. Northern Arizona University.
- Groom, N. 2005. 'Pattern and meaning across genres and disciplines: an exploratory study', *Journal of English for Academic Purposes* 4 (3), pp. 257–77.
- Halliday, M.A.K. 2004. *The Language of Science*. London: Continuum.
- Harwood, N. 2005a. "'I hoped to counteract the memory problem, but I made no impact whatsoever': discussing methods in computing science using *I*", *English for Specific Purposes* 24 (3), pp. 243–67.
- Harwood, N. 2005b. 'We do not seem to have a theory... the theory I present here attempts to fill this gap': inclusive and exclusive pronouns in academic writing', *Applied Linguistics* 26 (3), pp. 343–75.
- Hemais, B. 2001. 'The discourse of research and practice in marketing journals', *English for Specific Purposes* 20 (1), pp. 39–59.
- Hewings, M. and A. Hewings 2002. "'It is interesting to note that...': a comparative study of anticipatory 'it' in student and published writing", *English for Specific Purposes* 21 (4), pp. 367–83.
- Hyland, K. 1996. 'Writing without conviction? Hedging in science research articles', *Applied Linguistics* 17 (4), pp. 433–54.
- Hyland, K. 1999. 'Academic attribution: citation and the construction of disciplinary knowledge', *Applied Linguistics* 20 (3), pp. 341–67.
- Hyland, K. 2010. 'Constructing proximity: relating to readers in popular and professional science', *Journal of English for Academic Purposes* 9 (2), pp. 116–27.
- Hyland, K. and P. Tse. 2007. 'Is there an "academic vocabulary"?', *TESOL Quarterly* 4 (2), pp. 235–53.
- Koutsantoni, D. 2006. 'Rhetorical strategies in engineering research articles and research theses: advanced academic literacy and relations of power', *Journal of English for Academic Purposes* 5 (1), pp. 19–36.
- Martínez, I. 2005. 'Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English', *Journal of Second Language Writing* 14 (3), pp. 174–90.
- Peacock, M. 2006. 'A cross-disciplinary comparison of boosting in research articles', *Corpora* 1 (1), pp. 61–84.
- Tabachnick, B. and L. Fidell. 2007. *Using Multivariate Statistics*. (Fifth edition.) Boston, Massachusetts: Pearson.
- Warchal, K. 2010. 'Moulding interpersonal relations through conditional clauses: consensus-building strategies in written academic discourse', *Journal of English for Academic Purposes* 9 (2), pp. 140–50.

- Webber, P. 1994. 'The function of questions in different medical journal genres', *English for Specific Purposes* 13 (3), pp. 257–68.
- Wells, R. 1960. 'Nominal and verbal style' in T.A. Sebeok (ed.) *Style in Language*, pp. 213–20. Cambridge: Cambridge University Press.
- Williams, I. 1996. 'A contextual study of lexical verbs in two types of medical research report: clinical and experimental', *English for Specific Purposes* 15 (3), pp. 175–97.