

A multi-dimensional analysis of journal registers

7.1 Introduction

In the last two chapters, I focused on describing variation in the use of core grammatical features and a specific collection of features that function to ‘elaborate’ and ‘compress’ language. The study reported on in the present chapter takes a different approach, describing the disciplines and registers according to characteristic co-occurrence patterns in the use of a wide range of linguistic features. These linguistic features are not grouped into subsets based on discourse function (as was the case for the analysis of elaboration and compression) prior to the analysis. Rather, the analysis in this chapter uses the statistical method of factor analysis to locate patterns of linguistic features that statistically co-occur, an analytical approach to uncovering register variation cultivated and named multi-dimensional (MD) analysis by Biber (1988, 1995).

Factor analysis is a statistical method for data reduction. Tabachnick and Fidell (2007: 608) describe the goal of factor analysis: “to reduce a large number of observed variables to smaller number of factors”. That is, factor analysis serves to identify groups of variables that are correlated with one another (but not correlated with other groups of variables) in order to summarize trends in the data (see Tabachnick & Fidell 2007: 607–609). Biber (1988, 1995) applied statistical factor analysis to detailed linguistic analyses in order to characterize register variation in terms of the use of a much larger contingent of linguistic variables than previously examined in a single study, and to characterize variation that occurs along multiple parameters.

In Biber (1988), exploratory factor analysis was successfully used to identify patterns of variation across a wide range of spoken and written registers. The analysis resulted in the identification of seven factors, which were then interpreted as functional dimensions of variation. That is, each set of co-occurring features (and co-occurring features that were used in complementary distribution) were analyzed according to discourse functions generally common to the set of features included on a factor, with the interpretation aided by a consideration of

how the different registers related to those dimensions of variation (Conrad & Biber 2001: 24). In MD analysis, the term 'dimension' is used to encompass (typically) two groups of linguistic features, where the features within each respective group are highly correlated with the other variables in that group, and the two groups of variables occur in complementary distribution. In the past 20 years, MD analysis has been used in a variety of studies that consider the differences and similarities between highly diverse registers. In addition, MD analysis has been applied to the study of more specialized registers (e.g., see the chapters in Conrad & Biber 2001), and academic registers have been no exception. In the next section, I briefly summarize some of the research that has used MD analysis to study academic writing.

7.2 Background: Multi-dimensional analyses of academic language

As Conrad and Biber (2001) outline, MD analyses can be of two types. In the first type, registers are analyzed in terms of Biber's (1988) dimensions. In the second type, a full MD analysis is undertaken in which new dimensions are formulated based on the registers being investigated. Both approaches offer important information about register variation. When a new set of texts are assigned dimension scores for Biber's (1988) dimensions, they can be compared against the range of registers which have previously been analyzed in this way (as in Biber & Finegan 2001 and Conrad 1996a), adding to our knowledge about the range of registers that people encounter in their lives. The second approach, in which new dimensions are formulated, allows the researcher to discover co-occurring features that are important for a particular set of registers or genres.

In MD analyses of academic language, both types of analyses have been used. Conrad (1996b) uses Biber's (1988) dimensions to analyze textbooks and research articles in ecology and history, comparing them to other registers such as conversation, fiction, and non-fiction.¹ Conrad found that both register (textbooks versus research articles) and discipline corresponded to linguistic differences along the dimensions. For example, along Dimension 1, Conrad found that although research articles and textbooks in both disciplines were highly informational, research articles relied on those informational features a bit more. In contrast, she found that along Dimension 2 (narrative versus non-narrative discourse), variation

1. Conrad (1996a) also uses Biber's (1988) dimensions to analyze student texts; however, for the purpose of this book, I'll discuss only the results for professional academic texts here.

followed more closely along disciplinary lines, with Ecology registers being less narrative than history registers.

Biber and Finegan (2001) apply the MD approach to medical research articles, comparing the 1988 dimension scores for each of the IMRD (Intro-Method-Results-Discussion) sections to each other and to other registers. The analysis locates variation among the IMRD sections, but that variation is limited when compared to the range of variation found across other registers. For example, Biber and Finegan find that all sections of research articles are highly informational along Dimension 1 and generally non-narrative along Dimension 2. On Dimension 5 (impersonal versus non-impersonal), however, methods sections are very highly impersonal. Discussion, results, and introduction sections are also impersonal, but to a lesser extent than the methods sections.

Biber et al. (2004) apply the 1988 dimensions to a wider range of academic language, including both spoken and written registers encountered in university settings. Using the same corpora, Biber (2006: Chapter 7) conducts a new MD analysis to identify co-occurrence patterns that are specific to a more specialized domain (see Biber 2006: 181–182). Although the focus of this new analysis is on describing major register differences (e.g., between classroom teaching, textbooks, institutional writing, service encounters, etc.), Biber also includes a section in which he compares disciplines using classroom teaching and textbooks. If we look at the results specific to textbooks across disciplines, the analysis reveals that along the new Dimension 2 (procedural versus content-focused discourse), natural science textbooks are highly content-focused when compared to other disciplines (although all disciplines fall on the content-focused side of the continuum, see Biber 2006: 204). Biber found a wider range of variation across disciplines along his new Dimension 3 (narrative versus non-narrative orientation), with natural science and engineering textbooks being characterized as highly non-narrative, and education and humanities textbooks having a more narrative orientation.²

As Biber (2006: 181–182) notes, the MD approach is useful particularly when looking at many sub-corpora, such as varying registers in multiple disciplines. As summarized above, the previous MD analyses on academic language have been carried out on registers covering a broader range of situational characteristics, such as spoken and written academic registers. In the present study, the focus is on describing a much more specialized, narrow domain of language use. Perhaps more importantly, the focus is only on written registers. Thus, it is likely that the dimensions of variation found to be quite productive in describing

2. It should be noted, however, that classroom teaching in education, humanities, and social science relied on a narrative orientation to a much greater extent than these textbooks.

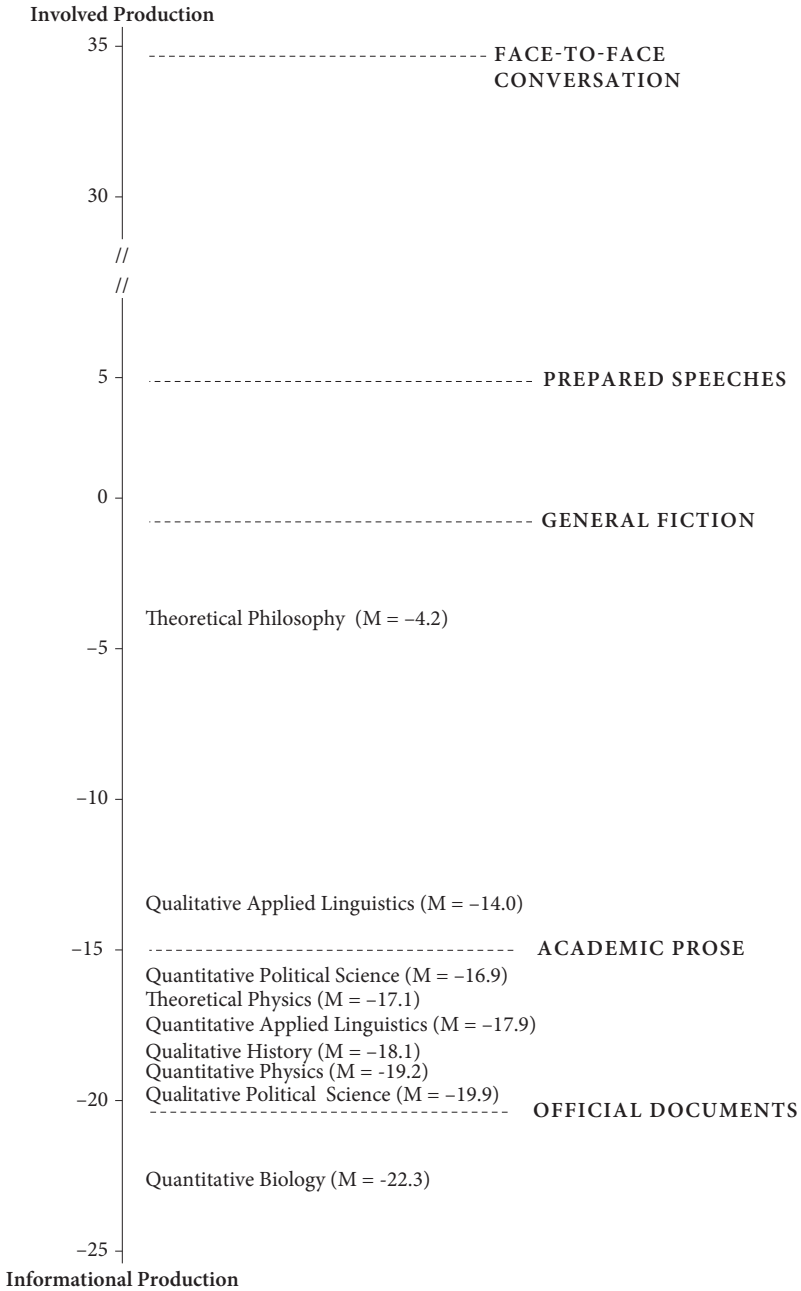


Figure 7.1. Distribution of disciplines and registers along Biber's (1988) Dimension 1 (involved versus informational production), with 5 general registers from Biber (1988) plotted for comparison

register differences in Biber (1988) will not be as useful in describing the range of variation that is important for these particular registers. In fact, we can readily see this by comparing the nine registers being studied in this book along the 1988 dimensions.

The tagcount program, described in Chapter 5, also produces dimension scores for each of the 1988 dimensions (dimension scores will be discussed below in Section 7.3.3). We can then plot these mean dimension scores by register and compare them to the registers studied in Biber (1988). Figure 7.1 shows this comparison along Dimension 1, which has been interpreted as characterizing 'involved' production versus 'informational' production and has represented a clear cline of variation that has consistently distinguished between many spoken and written registers, even across languages (e.g., see Biber 1995). In Figure 7.1, mean dimension scores for the 9 disciplines and registers in this study are plotted, showing that all of these registers fall on the 'informational' side, yet to differing extents. A selection of the registers from Biber (1988) are also plotted, showing that the range of variation among these registers is much broader, while the range of variation among the disciplines and academic journal registers is much more restricted in nature.

Thus, while making comparisons such as this one are interesting in and of themselves, another approach is to conduct a new factor analysis in order to identify the patterns of variation that are the most important in the more specialized domain of types of journal articles across disciplines. In the next section, I describe the methodology for carrying out a new multi-dimensional analysis of academic journal registers across disciplines.

7.3 Carrying out a new multi-dimensional analysis

7.3.1 Initial factor analyses to determine linguistic variables

One of the major strengths of multi-dimensional analyses of language variation is the ability to at once consider the use of very large sets of linguistic features. The data for this MD analysis comes from the 'tagcount' program described briefly in Chapter 3, which provides counts for approximately 130 linguistic features ranging from lexical classes to syntactic structures (see Conrad & Biber 2001: Chapter 2 for a detailed overview).

The first step in a new MD analysis is to select the linguistic features that will be used in the factor analysis. To be suitable for inclusion in a factor analysis, linguistic variables should be conceptually distinct. That is, the linguistic features should be correlated to some degree in terms of use, but care should be

taken that variables largely measuring the same construct or linguistic feature are not all included in the factor analysis. A competing goal is to be as inclusive and specific as possible in the selection of variables for the analysis (Conrad & Biber 2001: 15). For example, the output from the tagcount program includes counts for the overall use of *that*-complement clauses controlled by verbs as well as separate counts for *that*-clauses controlled by stance-carrying verbs grouped into different meaning subsets, such as non-factive, factive, attitudinal, and likelihood verbs. The first count is in essence a composite variable made up of the individual counts for specific subsets of verbs, and thus these variables are not conceptually distinct. Therefore, a choice has to be made whether to include a greater number of more specific variables (one for each type of stance meaning), or a more general overall count. This decision is made by carrying out pilot factor analyses to determine which combinations of variables are able to explain the greatest proportion of variance in the corpus.

In order to determine the set of variables to use in the final factor analysis, a series of initial factor analyses were run using varying groupings of the possible variables. These initial runs served as pilot analyses to identify features which were likely to contribute to the explanation of the linguistic patterns. The various pilot factor analyses were compared in terms of the linguistic features that 'loaded' onto the factors, along with the amount of total variance in the corpus that is accounted for by the various pilot solutions. In general, linguistic features were included in the final analyses only if their communalities exceeded .250 and they loaded on at least one factor with a factor loading³ of greater than .30.⁴ Based on the cumulative patterns found in these pilot factor analyses, 70 features were selected to be included in the final factor analysis, listed in Table 7.1. All words in the semantic sets of nouns, verbs and adjectives are listed in Appendix C.

3. Factor loadings range from 0 to 1, and measure how much variance an individual feature has in common with the total shared variance of a factor, thus indicating degree of co-occurrence between feature and the set of features on the factor (Conrad & Biber 2001: 21; Biber 1988).

4. Communalities represent the amount of variance in that variable that is explained by the factor solution (Tabachnick & Fidell 2007: 621). Variables with low communalities are excluded from the factor analysis.

Table 7.1. Summary of linguistic features included in the final factor analysis

Linguistic Feature	Description/Example
<i>A. General</i>	
1. type-token ratio	in the first 400 words of the text
2. word length	average number of letters per word
3. word count	overall number of words per text
<i>B. Nouns and Pronouns</i>	
4. 1st person pronouns	e.g., <i>I, we</i>
5. 2nd person pronouns	e.g., <i>you</i>
6. 3rd person pronouns	e.g., <i>he, she, they</i>
7. pronoun 'it'	all instances of 'it'
8. demonstrative pronouns	<i>this, these, that, those</i>
9. nominal pronouns	e.g., <i>somebody, anyone</i>
10. all nouns	all words identified as nouns by automatic tagger
11. nominalizations	e.g., <i>interaction, communication</i>
12. animate nouns	e.g., <i>adult, applicant, child, immigrant, patient</i>
13. process nouns	e.g., <i>achievement, comparison, effect, formation</i>
14. cognition nouns	e.g., <i>ability, decision, concept, idea, knowledge</i>
15. other abstract nouns	e.g., <i>advantage, background, culture, model</i>
16. concrete nouns	e.g., <i>acid, brain, camera, computer, glacier</i>
17. technical nouns	e.g., <i>atom, cell, compound, equation, message</i>
18. quantity nouns	e.g., <i>amount, century, frequency, percentage</i>
19. group nouns	e.g., <i>church, committee, government, institute</i>
<i>C. Verbs</i>	
20. possibility, permission and ability modals	<i>can, could, may, might</i>
21. prediction modals	<i>will, would, shall, be going to</i>
22. necessity and obligation modals	<i>must, should, had better, have to, got to, ought</i>
23. verb <i>BE</i>	all forms of verb <i>BE</i>
24. verb <i>HAVE</i>	all forms of verb <i>HAVE</i>
25. activity verbs	e.g., <i>bring, combine, encounter, obtain, produce</i>
26. communication verbs	e.g., <i>acknowledge, answer, claim, discuss</i>
27. mental verbs	e.g., <i>confirm, find, identify, observe, predict, think</i>
28. causative verbs	e.g., <i>affect, allow, help, influence, require</i>
29. existence verbs	e.g., <i>appear, define, illustrate, indicate, reflect</i>
30. aspectual verbs	e.g., <i>begin, complete, continue, keep, start</i>

(Continued)

Table 7.1. (Continued) Summary of linguistic features included in the final factor analysis

Linguistic Feature	Description/Example
<i>D. The Verb Phrase</i>	
31. past tense	e.g., <i>claimed, concluded, found, reported</i>
32. perfect aspect	e.g., <i>had argued, have discussed, has shown</i>
33. progressive aspect	e.g., <i>is becoming, is causing, are seeking</i>
34. agentless passive voice	passive constructions with no specified agent
35. by-phrase passive voice	passive constructions with agent in by-phrase
<i>E. Adjectives</i>	
36. all attributive adjectives	all adjectives occurring as a noun pre-modifier
37. all predicative adjectives	all adjectives occurring in post-predicate position
38. size adjectives (attributive)	e.g., <i>big, great, large, small</i>
39. time adjectives (attributive)	e.g., <i>new, young, old</i>
40. evaluative adjectives (attributive)	e.g., <i>best, good, important</i>
41. relational adjectives (attributive)	e.g., <i>basic, common, different, major, similar</i>
42. topical adjectives (attributive)	e.g., <i>economic, human, international, public</i>
<i>F. Adverbs</i>	
43. general adverbs	
44. time adverbs	e.g., <i>again, later, now</i>
45. stance adverbs	e.g., <i>obviously, evidently, frankly, surprisingly</i>
<i>G. Coordination and Subordination</i>	
46. adverbial conjuncts	e.g., <i>however, therefore, thus</i>
47. clausal coordinating conjunctions	e.g., <i>and, or</i>
48. phrasal coordinating conjunctions	e.g., <i>but</i>
49. conditional subordinating conjunctions	e.g., <i>if, unless</i>
50. subordinating conjunctions (other)	e.g., <i>as, except</i>
<i>H. Clauses Marking Stance</i>	
51. <i>that</i> -clause controlled by non-factive (communication) verb	e.g., <i>argue, claim, show, tell</i>
52. <i>that</i> -clause controlled by factive (certainty) verb	e.g., <i>demonstrate, conclude</i>
53. <i>that</i> -clause controlled by likelihood verb	e.g., <i>appear, estimate, seem, suppose, suggest</i>
54. <i>that</i> -clause controlled by factive (certainty) adjective	e.g., <i>conclude, proves</i>
55. <i>that</i> -clause controlled by likelihood adjective	e.g., <i>possible, probable</i>

(Continued)

Table 7.1. (Continued)

Linguistic Feature	Description/Example
<i>H. Clauses Marking Stance</i>	
56. <i>that</i> -clause controlled by attitudinal adjective	e.g., <i>afraid, aware, surprised</i>
57. <i>that</i> -clause controlled by factive (certainty) noun	e.g., <i>conclusion, fact, observation</i>
58. <i>that</i> -clause controlled by likelihood noun	e.g., <i>assumption, belief, hypothesis</i>
59. <i>that</i> -clause controlled by attitudinal noun	e.g., <i>hope, fear, view</i>
60. <i>to</i> -clause controlled by speech verb	e.g., <i>ask, claim, show</i>
61. <i>to</i> -clause controlled by verb of desire	e.g., <i>agree, hope, intent, prefer</i>
62. <i>to</i> -clause controlled by verb of causality, modality	e.g., <i>attempt, help, permit, require</i>
63. <i>to</i> -clause controlled by verb of probability	e.g., <i>appear, seem, tend</i>
64. all <i>to</i> -clauses controlled by stance adjectives	e.g., <i>certain, worried, appropriate, difficult, easy</i>
65. all <i>to</i> -clauses controlled by stance nouns	e.g., <i>claim, possibility, assumption, fact</i>
<i>I. Post-Nominal Modifiers</i>	
66. passive postnominal modifier	non-finite -ed clause postmodifying a noun
67. <i>that</i> relative clause	relative clause with <i>that</i> as relative pronoun
<i>J. Other</i>	
68. <i>wh</i> -questions	all clauses tagged as <i>wh</i> -questions
69. <i>wh</i> -clauses	all clauses with <i>wh</i> -complementizer
70. all prepositions	any word tagged as a preposition

7.3.2 Final factor analysis

Two preliminary analyses were used to ensure that the data set of selected variables is appropriate for factor analysis (using SPSS v. 19.0). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO = .855, meritorious) meets the minimum requirement for FA (values greater than .6, Tabachnick & Fidell 2007). Likewise, Bartlett's Test for Sphericity (Approximate Chi-Square = 11005.23, df = 2415, $p = .000$) is significant, indicating that the null hypothesis that correlations are 0 can be rejected (Tabachnick & Fidell 2007). Both of these tests indicate that adequate correlations exist in the correlation matrix, and thus, FA is suitable for the data.

The final factor analysis was carried out using Principal Axis Factoring with Promax rotation in IBM SPSS v19.0. The four-factor solution was determined to be the most interpretable when compared with three- and five-factor solutions. Appendix D lists the full (rotated) factorial structure of the four-factor solution, and Appendix E shows the scree plot for the solution. The Initial Eigenvalues (Total Variance Explained) indicate a cumulative percentage of variance explained as 40.26%. All variables with a factor loading of .30 or above were considered important for the analysis for that factor. One variable (*that*-clauses controlled by attitudinal adjectives) did not load on any factor at the specified level.

7.3.3 Calculating and comparing factor scores across disciplines and registers

After identifying the underlying patterns of variation using factor analysis, the next step in the analysis is to characterize each register according to the factors themselves, that is, to quantify the extent to which each register utilizes the patterns of variation uncovered by the factor analysis. First, factor scores were calculated for each dimension for each text in the corpus. To calculate factor scores, z-scores were first computed for each linguistic feature in order to convert the rate of occurrence for each variable into a standardized scale where the mean is equal to 0 and the standard deviation is 1. This serves to equalize the impact of high- and low-frequency variables, so that dimension scores are not disproportionately impacted by high frequency linguistic features (see Biber 1988: 94).

Each factor in this study resulted in two groupings of variables, 'positive' features and 'negative' features, which represent co-occurring sets of features that are in complementary distribution. In other words, texts which rely highly on the positive features rely on the negative features to a lesser extent, and vice versa. To calculate dimension scores, which indicate the degree to which a text can be said to rely on linguistic features on a factor, the standardized z-scores for all of the positive features are added together, and the z-scores for the negative features are subtracted (Biber 1988).

Once a factor score has been computed for each text in the corpus for each dimension, dimension scores for the disciplines/registers can be calculated by taking the mean factor score for each text in sub-corpus. The means and standard deviations for each dimension are included in Appendix F, and Figures 7.2 – 7.5 below plot the registers according to mean dimension scores. One-way ANOVAs were used to test the significance of each dimension (Appendix F, Table F2), and show that all four factors are significant at $\alpha < .05$. The ANOVA results, along with an r^2 value, are presented along with Figures 7.2 – 7.5. The r^2 value is a measure of the proportion of the variance in the dimension scores that can be explained by the register groupings of the texts, and thus indicates how important a dimension

is for explaining the variation in the corpus (see Conrad & Biber 2001:28). The r^2 values for the four dimensions in this MD analysis range from .81 (Dimension 2) to .43 (Dimension 4).

To test for individual differences between disciplines and registers, the Games-Howell procedure (equal variances not assumed, see Appendix F, Table F4) was used to make post-hoc comparisons. These post-hoc comparisons are listed in Appendix F, Tables F5 – F8. In the next sections, I turn to the interpretation of the four factors, relying cyclically on functional interpretations of the linguistic features that characterize a dimension of variation, as well as a consideration of how the disciplines and registers fall along these parameters.

7.4 Dimensions of variation in academic journal registers in 6 disciplines

Table 7.2 summarizes the four factors, listing the sets of co-occurring linguistic features with factor loadings indicated in parentheses. I have also given each factor a descriptive title that previews the functional analysis of these dimensions:

Dimension 1: Academic Involvement & Elaboration vs. Information Density

Dimension 2: Contextualized Narration vs. Procedural Description

Dimension 3: Human Focus vs. Non-Human Focus

Dimension 4: ‘Academese’

In the sections that follow, I explore each of dimensions of variation in detail. More specifically, I discuss the functional underpinnings of these groups of features in relation to the ways in which the disciplines and registers are distributed along these dimensions of variation, and illustrate this analysis with text excerpts throughout.

7.4.1 Dimension 1: Academic involvement and elaboration vs. informational density

Dimension 1, labeled academic involvement and elaboration versus informational density, is made up of 26 features on the positive end of the factor, and 8 features on the negative end of the factor. Despite the fact that this dimension has been extracted based on only written registers in a fairly specialized domain, there is a good deal of overlap in the features that comprise this factor, Biber’s (1988) Dimension 1 (involved versus informational discourse) and Biber’s (2006) Dimension 1 (oral versus literate texts). In fact, this overlap is particularly apparent for the negative features on this dimension, where all but two of the features (process nouns, past tense) were also negative features in Biber’s (1988) and/or (2006) dimensions:

Table 7.2. Structure of four-factor solution**Dimension 1: Academic Involvement & Elaboration vs. Information Density***Positive features:*

Pronouns: nominal pronouns (.69), pronoun *it* (.62), 1st person pronouns (.58), demonstrative pronouns (.52)

Nouns: nouns of cognition (.57)

Adjectives: predicative adjectives (.70), evaluative attributive adjectives (.33)

Verbs: verb *be* (.79), verb *have* (.67), causative verbs (.34)

Modal Verbs: modals of prediction (.69), modals of possibility (.66), modals of necessity (.65)

Adverbs: general adverbs (.54), stance adverbials (.47), adverbials of time (.34)

Conjunctions: Subordinating conjunction – conditional (.83), adverbial conjuncts (.48), subordinating conjunctions (.39)

Finite Clauses: *that*-clauses controlled by nouns of likelihood (.65), *that*-clauses controlled by verbs of likelihood (.59), *that*-clauses controlled by factive adjectives (.48), *that*-clauses controlled by attitudinal nouns (.47), *that*-clauses controlled by factive nouns (.44), *wh*-clauses (.34)

Non-Finite Clauses: *to*-clauses controlled by stance adjectives (.37), *to*-clauses controlled by verbs of probability

Negative features:

Nouns: nouns (–.75), process nouns (–.40)

Verbs: past tense verbs (–.67)

Passives: passive postnominal modifiers (–.53), agentless passive voice verbs (–.32)

Other: prepositions (–.39), type-token ratio (–.35), word length (–.31)

Dimension 2: Contextualized Narration vs. Procedural Description*Positive features:*

Pronouns: 3rd person pronouns (.65)

Nouns: group nouns (.49), nominalizations (.32), animate nouns (.43)

Adjectives: topical attributive adjectives (.53), attributive adjectives indicating time (.47)

Verbs: past tense verbs (.55), aspectual verbs (.52), perfect aspect verbs (.48), communication verbs (.47), present progressive verbs (.42)

Conjunctions: phrasal coordinating conjunctions (.51), clausal coordinating conjunctions (.35)

Finite Clauses: *that*-relative clauses (.46), *that*-clauses controlled by non-factive verbs (.45), *wh*-questions (.32)

Non-Finite Clauses: *to*-clauses controlled by verbs of modality, causation and effort (.57), *to*-clauses controlled by verbs of desire (.41), *to*-clauses controlled by stance nouns (.35)

Other: word length (.52), word count (.36), type-token ratio (.31)

(Continued)

Table 7.2. (Continued)

*Negative features:***Nouns:** technical nouns (–.61), quantity nouns (–.46), concrete nouns (–.37)**Adjectives:** attributive adjectives indicating size (–.37)**Passives:** agentless passive voice verbs (–.52), passive voice verbs with *by*-phrases (–.47)**Dimension 3: Human vs. Non-human Focus***Positive features:***Pronouns:** 2nd person pronouns (.40), 3rd person pronouns (.35)**Noun:** process nouns (.50)**Verbs:** mental verbs (.65), activity verbs (.60), communication verbs (.51), present progressive verbs (.49)**Finite Clauses:** *that*-clauses controlled by factive verbs (.42), *wh*-clauses (.33)**Non-Finite Clauses:** *to*-clauses controlled by verbs of desire (.46), *to*-clauses controlled by speech verbs (.45)*Negative features:***Adjectives:** attributive adjectives (–.54), attributive adjectives indicating topic (–.42)**Adverbs:** general adverbs (–.31)**Other:** prepositions (–.35)**Dimension 4: ‘Academese’***Positive features:***Nouns:** nominalizations (.43), process nouns (.38), other abstract nouns (.32)**Adjectives:** relational attributive adjectives (.45)**Verbs:** existence verbs (.37)**Finite Clauses:** *that*-clauses controlled by likelihood adjectives (.38), *to*-clauses controlled by stance adjectives (.33)**Other:** word length (.58)*Negative features:***Adverbs:** time adverbials (–.47)

all nouns, passive post-nominal modifiers, agentless passive verbs, prepositions, type-token ratios, and word length. In previous multi-dimensional research on a range of registers, these features have been associated with informational purposes, particularly in written registers. It is interesting that this combination of features still emerges as a dimension of variation that distinguishes amongst written texts which all have a primary informational purpose. This finding means that we must take a deeper look at how these features are functioning in the particular

registers in the study. When we do so (illustrated below), we see that these features can also be associated more specifically with *informational density* – that is, with highly compressed styles of discourse in which a great deal of information is presented in a dense manner.

On the positive end of the factor, features include various types of pronouns (nominal pronouns, ‘it’, 1st person pronouns, and demonstrative pronouns), modifiers (predicative adjectives, evaluative attributive adjectives, general adverbs, adverbs of time), explicit markers of logical and grammatical relationships (conditional and other subordinating conjunctions, adverbial conjuncts), and structures that convey personal stance meanings (possibility/permission/ability modals, *that*-clauses controlled by likelihood nouns, verbs and adjectives, factive nouns and adjectives, and attitudinal nouns; *to*-clauses controlled by stance adjectives and verbs of probability).

Although the degree of correspondence between the positive features on this dimension and Biber’s (1988, 2006) dimensions is not as strong as for the negative features, some functional overlap does exist. For example, Biber’s (1988) dimensions also contained many stance markers (including emphatics, hedges, and amplifiers),⁵ types of pronouns, and explicit grammatical links (e.g., causative subordination and non-phrasal coordination, see Biber 1988). Further overlap exists between Biber’s (2006) Dimension 1, including some of these same features, as well as conditional and *wh*-clauses.

In contrast, many of the features which were important on Biber’s earlier dimensions, but which are not important features along Dimension 1 in the present study, are those that are highly correlated with interactional spoken language: contractions, indefinite pronouns, *wh*-questions, *that*-deletion, and discourse particles (see Biber 1988, 2006; Friginal 2009). In fact, many of these features were simply not included in the factor analysis conducted in this study because they did not exhibit high frequencies or substantial variation across these registers during the initial pilot factor analyses. That is, they did not contribute to understanding linguistic variation in these registers (as they are primarily characteristics of spoken language) and were excluded from further analysis.

Taking into account these two sets of features, it appears that Dimension 1 in this study reflects a common distinction that has been consistently identified in several different multi-dimensional analyses: that of high-density informational language and more involved language production. It should be noted, however,

5. In fact, these linguistic features from the (1988) dimensions were not considered in the present factor analysis, as these categories were largely included in various other stance categories from the newer stance framework developed based on the LGSWE (Biber et al. 1999), which has been used in more recent studies.

that Dimension 1 here reveals a narrower scope of involvement, more focused on those involvement features that are key to academic writing. Support for this interpretation is apparent if we look at the distribution of the disciplines and registers in the present study, plotted by mean dimension score in Figure 7.2. While not an exact match, the general pattern of how registers fall along Dimension 1 in the present study is quite similar to how they fall along Biber's (1988) Dimension 1, displayed above in Figure 7.1. Most noticeably, Figure 7.2 shows that Dimension 1 reflects a dichotomy between philosophy and all other disciplines. Philosophy has an *extremely* high positive dimension score (41.7) while the other disciplines range from a low of -15.9 (biology) to 2.1 (quantitative political science). This is the same overall pattern as seen above in Figure 7.1. The new Dimension 1 thus appears to be reflecting the same general underlying pattern, yet results in more precise information as to the involvement structures that are playing important roles in academic writing specifically (or at least in theoretical philosophy).

The issue of compression in academic language has been addressed to a certain extent in the last chapter; however, the statistical co-occurrence of some of these same compression features confirms the importance of these densification structures for academic writing. Excerpt 7.1 illustrates the high information density, where various structures are embedded into complex noun phrases. The last half of this excerpt is particularly reflective of the dense information structure that results from the use of the negative features on this dimension, where appositive noun phrases are used in abundance to pack concepts into a sentence structure with relatively few verbs. In fact, in the passage of 161 words, only five main clause verbs phrases (*were used*, *were delimited*, *were not included*, *were collected*, and *were collected*) are used, all of them passive. In the excerpt, the negative features on Dimension 1 are marked: nouns are **bolded**, prepositions are underlined, past tense verbs are *italicized*, passive structures are double-underlined, and passive postnominal modifiers are additionally in SMALL CAPS.

- 7.1 *Quantitative Biology* (Hoeinghaus, Winemiller, & Agostinho 2008):
 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**. **Parasites** and non-aquatic **organisms** that feed on fish, such as **birds** and **humans**, *were not included*. **Samples** for isotopic analysis were collected between **September** and early **December** 2003 (late dry season), prior to seasonally rising **water levels** and **fish migrations**. At each **location**, representative riparian and aquatic **carbon sources** (**C plants** and **C grasses**, fine **particulate organic material**, coarse **detritus**, **periphyton**, and **phytoplankton**), primary **consumers**

ACADEMIC INVOLVEMENT & ELABORATION

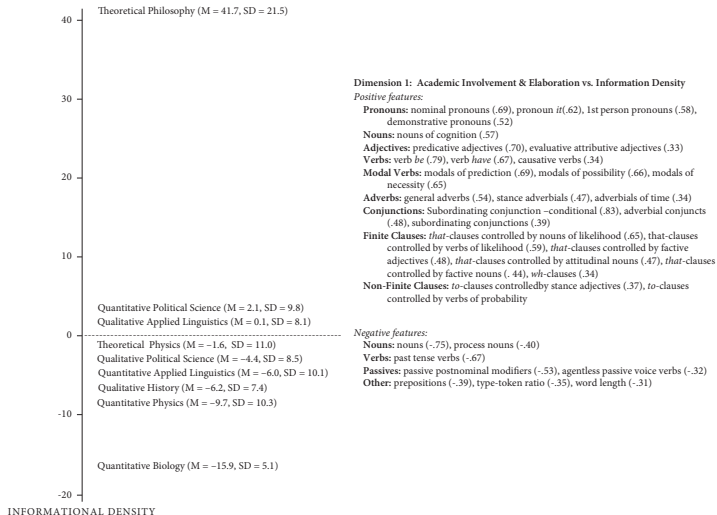


Figure 7.2. Distribution of disciplines and registers along Dimension 1: Academic involvement and elaboration versus informational density. One-way ANOVA results: $F = 66.62$, $p = .000$, $r^2 = .67$. Post-hoc comparisons are listed in Appendix F

(snails, bivalves, zooplankton, and herbivorous and detritivorous fishes) and secondary consumers (omnivorous and carnivorous fishes) were collected at multiple points along a 2A5 km sample reach to characterize trophic pathways from source to top consumer.

While the excerpt in 7.1 shows almost no use of the positive features on Dimension 1, it is not the case that registers with highly negative dimension scores on Dimension 1 lack all positive features. Rather, if we mark another passage from quantitative biology for both the negative and positive features on Dimension 1, we can see that the use of the positive features is simply overshadowed by the density of the negative features. In the first mark-up of this excerpt, the negative features of Dimension 1 are marked: nouns are **bolded**, prepositions are underlined, past tense verbs are *italicized*, passive structures are double-underlined, and passive postnominal modifiers are additionally in SMALL CAPS.

7.2 *Quantitative Biology (Kelly, Macisaac, & Heath 2006):*

The beam waist at 7 mm along the beam from the minimum waist is increased only by 10%. This length of the beam is within the main viewing angle of the detector. 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). This means also systematically, that we do not use the disproved **models** to attempt to interpret later **steps in the development**, to avoid **confusion**.

In the second mark-up of this same excerpt, the positive features are marked: (main clause verbs are in bolded **SMALL CAPS**, pronouns are underlined, general adverbs and predicative adjectives are underlined, prediction and possibility modal verbs are *italicized*, and clausal structure are indicated with head words **bolded** and corresponding [square brackets]).

7.3 *Quantitative Biology (Kelly, Macisaac, & Heath 2006):*

The beam waist at 7 mm along the beam from the minimum waist is **INCREASED only** by 10%. This length of the beam **IS** within the main viewing angle of the detector. 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 if₁ [they might have been proved previously in other publications (always referenced).]₁ **This MEANS₂** also systematically, [that we do not use the disproved models to attempt to interpret later steps in the development, to avoid confusion.]₂

The first excerpt from quantitative biology (excerpt 7.1), which focused on describing research procedures, used very few positive features along Dimension 1. In contrast, the passage in 7.2 and 7.3 exhibits some instances of positive features along Dimension 1. However, in addition to illustrating the (still) much less frequent use of positive features, this passage also demonstrates that in biology the positive features on Dimension 1 are used at points in the discourse where the writer is commenting on the research, making connections outside of the actual events of the research, and discussing the implications of the research. This contrasts to the use of the positive features of Dimension 1 in philosophy (excerpt 7.4 below), where the features are relied upon more pervasively throughout the discourse.

In the philosophy excerpt below, the positive features on Dimension 1 seem to be used for two main purposes. First, as in Biber's earlier dimensions, the use of personal pronouns creates a sense of interaction in the discourse, with authors explicitly talking about themselves (as in the biology example above), and bringing others into the study in order to illustrate points or provide evidence. Second, the features on Dimension 1 are used to elaborate discourse and show the involvement of the writers by providing personal stance and evaluation. These explicit evaluations, along with other types of subordinating conjunctions, serve to make explicit connections between meaning relationships, interpretations, and the authors' own evaluations, as illustrated in excerpt 7.4 (The positive features are again marked as above, and additional positive features such as cognition nouns and *wh*-clauses are underlined):

7.4 *Theoretical Philosophy (Roache 2006):*

Wollheim's **point**₁ [that in order to have q-memories of everything that his father experienced on his childhood walks, he *must* inherit not only his father's memories of these walks, but also their psychological context?]₁ Well, it *may* prove a problem for Wollheim's **claim**₂ [that it *is* impossible₃ [to isolate memories from their psychological context,]₃]₂ but it does not undermine our weaker **claim**₄ [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 It seems generally true₈ [that a loss of psychological context results in an impoverishment in the content of a memory.]₈ [PHIL-TH]

In contrast, we can look at this same passage with the negative features highlighted in order to illustrate the less frequent use of the negative features of Dimension 1: nouns are **bolded**, prepositions are underlined, passive structures are double-underlined, and passive postnominal modifiers are additionally in SMALL CAPS.

7.5 *Theoretical Philosophy (Roache 2006):*

Wollheim's point that in order to have **q-memories** of everything that his father experienced on his **childhood walks**, he must inherit not only his **father's memories** of these walks, but also their psychological **context**? Well, it may prove a **problem** for Wollheim's claim that it is impossible to isolate **memories** from their psychological **context**, but it does not undermine our weaker claim that **memories** are weakened when they are isolated in this way. Martin does not tell us how much of the Spanish **conversations** he now remembers, but it seems plausible to suppose that the **fact** that he no longer understands **Spanish** will have resulted in his now not being able to recall some of the **details** of the **conversations**. It seems generally true that a **loss** of psychological **context** results in an **impoverishment** in the **content** of a **memory**. [PHIL-TH]

Biology and philosophy illustrate the two extremes of Dimension 1, but Figure 7.2 shows less variation for the remaining disciplines and registers. While quantitative physics is also highly negative, the remaining disciplines (applied linguistics, political science, history, and theoretical physics) all have mean dimension scores that fall within an approximately 6 point range near 0. This clustering of disciplines and registers around 0 on the Dimension 1 scale shows that these disciplines and registers rely on the elaborating features and the information density features in more balanced ways than philosophy and biology. We can see this more balanced use of elaborated and densification features in the following excerpts. Here, for the sake of clarity, all positive features along Dimension 1 are **bolded**, while all negative features are underlined. In excerpt 7.6, we can see the overall frequent use of nouns and prepositions that help maintain a nominal style of writing, while we can also see cognition nouns (e.g., *theories*, *concepts*), *be* verbs, predicative adjectives (e.g., *are useful*, *is crucial*), modal verbs, general adverbs, adverbial conjuncts (e.g., *however*) and stance to-clauses (e.g., *crucial to identify X*).

7.6 *Qualitative Applied Linguistics (Frazier 2007):*

A large number of studies investigate the talk of students in writing classrooms; most of these treat the act of writing as social in nature. 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 to identify the interactional details of group work discourse. Most of the sources that investigate writing students' talk (some of which are 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.

In theoretical physics, we see the same dense use of nouns and prepositions, past tense verbs (e.g., *used*, *obtained*). However, we also see positive features such as adverbial conjuncts (e.g., *however*, *in addition*), personal and demonstrative pronouns, *be* as main verb, and several *that*-complement clauses. In this excerpt, previous research is being discussed and connected to the issues relevant for the study to be reported on.

7.7 *Theoretical Physics (Thomas, Christakis, & Jorgensen 2006):*

However, Menger and D'Angelo used ¹³C NMR to observe the conformational equilibria of undecane-2,5-di-¹³C in solvents ranging from chloroform to aqueous ethanol; by measuring 3JCC, **they** obtained the fraction of trans and gauche for the C–C bond. The result was 76% trans in all solvents. **In addition**, small-angle neutron scattering (SANS) studies by Dettenmaier and GoodsaidZalduondo and Engelman and Raman studies by Fischer agree [that the influence of intermolecular interactions on individual monomer conformation in the liquid state] is **negligible** and [that the conformations of n-alkanes in the condensed phase are similar to those populated in isolation.]

As mentioned above, the distribution of the registers along Dimension 1 show a primarily dichotomous relationship. This dichotomy is further supported by the post-hoc comparisons, listed in Appendix F. These comparisons show that philosophy and quantitative biology are significantly different from nearly every other register (with one exception: biology and quantitative physics are not significantly different). The remaining post-hoc comparisons reveal only a few other significant differences (e.g., quantitative physics is significantly different from quantitative political science and qualitative applied linguistics).

However, the placement of two registers is somewhat surprising: theoretical physics and qualitative history. Looking past the dichotomy between theoretical philosophy and the other disciplines, at first glance it appears that there is also a distinction between the natural sciences and the remaining non-science disciplines, with quantitative biology and quantitative physics having the largest negative scores along Dimension 1. However, we then notice that theoretical physics has a mean dimension score near zero, while three non-science registers have more highly negative scores. Particularly surprising here is qualitative history, with the lowest negative score of the non-science disciplines (-6.2). I'll return to history in a moment, but first let's take a look at theoretical physics.

Excerpt 7.7 illustrated how the positive and negative features were used in combination during a discussion of previous research and its relation to the upcoming study. However, like theoretical philosophy, theoretical physics relies on the logical progressions of evidence in order to support claims and present

research, and Excerpt 7.8 employs positive features (all **bolded**) such as 1st person pronouns, modal verbs, time adverbs, and demonstrative pronouns in order to involve and guide the reader through the steps in the analyses. However, the use of nouns, prepositions, and passive verbs are also still prevalent in the excerpt.

7.8 *Theoretical Physics (Cacciatori et al. 2008):*

The second constraint, in particular Eq. (3), **can** be used to restrict the search for the [formula] functions [formula] **to that** of a single one, for which **we** choose [formual] with [formula]. **We** work out the details of this reduction. In particular, **we** give the constraints which the function [formual] **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 $\text{Sp}(2g, \mathbb{Z})$ 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.

While it was a bit unexpected to see the use of positive features along Dimension 1 in physics, it was also surprising to see the frequent use of negative features in qualitative history. In the following excerpt, negative features along Dimension 1 are marked: nouns are **bolded**, prepositions are underlined, past tense verbs are *italicized*, passive structures are double-underlined, and passive postnominal modifiers are additionally in SMALL CAPS.

7.9 *Qualitative History (Lopes Don 2006):*

By the early 1540s, a **consensus** *developed* in **councils** of the Spanish **government** that the use of the **Inquisition** to induce religious **orthodoxy** among the new **converts** *was* inappropriate and possibly dangerous for the **security** of the colony. 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**.

This excerpt is typical of qualitative history reports, showing a dense use of nouns and prepositions, many of which function as noun modifiers and result in compressed noun phrase structures. However, unlike the natural sciences, passive constructions are not notably frequent (also documented in Chapter 5). Rather, the use of the past tense is extremely frequent as past events are described and reported. Combining with this frequent use of nouns, prepositions, and past tense

verbs is the relative lack of many positive features along Dimension 1, such as modal verbs and stance-conveying constructions – linguistic features not used frequently in the reporting of the events and states of the past.

In sum, Dimension 1 illustrates (A) a dichotomous division between philosophy and other registers, (B) a hint of a pattern in which empirical, natural science research relies on features related to informational density to a greater extent than social science research, and (C) that individual registers and disciplines exhibit somewhat idiosyncratic patterns of use based on situational characteristics unique to that discipline or discipline/register combination. In the following section, however, we see a different type of pattern of variation.

7.4.2 Dimension 2: Contextualized narration vs. procedural discourse

A total of 28 features have factor loadings greater than or equal to .30 on Dimension 2, with 22 positive features and 6 negative features. Positive features include word classes that refer to human constituents (3rd person pronouns, group nouns, animate nouns), adjectives indicating time and topic, tense and aspect markers (past tense, perfect and progressive aspect verbs), *that*-relative clauses, and communication verbs (including *that*-clauses controlled by non-factive/communication verbs). In addition, phrasal and clausal coordinating conjunctions loaded with the positive features on Dimension 2. Negative features on Dimension 2 include three specific types of nouns (technical nouns, quantity nouns, and concrete nouns), attributive adjectives indicating size, and agentless and by-phrase passive voice verbs.

Figure 7.3 plots the mean dimension scores by register along Dimension 2. In contrast to Dimension 1, where the nature of the discipline reflected the two poles of the variation, a different organization of registers emerges for Dimension 2. The three qualitative registers (history, political science, and applied linguistics) have the highest positive dimension scores, followed by theoretical philosophy. The hard sciences, quantitative biology and quantitative and theoretical physics, have highly negative dimension scores, while the quantitative registers in the social sciences (political science and applied linguistics) fall in the middle of this dimension. As I will explore in the rest of this section, it appears that Dimension 2 is highly correlated with the primary way in which the disciplines and registers present evidence. That is, the three qualitative registers and theoretical philosophy all rely upon extensive prose to present evidence, and this prose often provides a narration of what happened during a research study with participants (i.e., qualitative applied linguistics) or during a particular time period or political situation (i.e., history and political science).

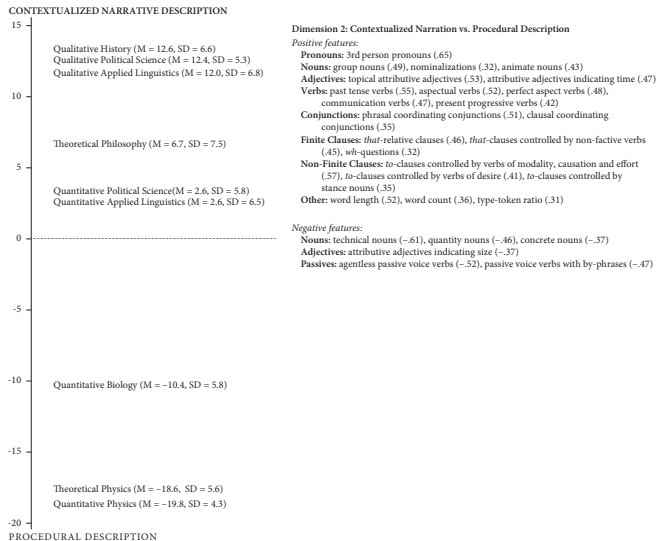


Figure 7.3. Distribution of disciplines and registers along Dimension 2: Contextualized narration versus procedural description. One-way ANOVA results: $F = 138.99$, $p = .000$, $r^2 = .81$. Post-hoc comparisons are listed in Appendix F

In fact, many of the positive features have been linked in the past to narrative discourse (e.g., see Biber 1988: Dimension 2; Biber 2006: Dimension 3), particularly 3rd person pronouns, human/animate nouns, past tense and perfect aspect verbs, and communication verbs. This interpretation also fits with the way in which these features are used in the three qualitative sub-corpora. In the following excerpts, 3rd person pronouns, group nouns, and animate nouns are underlined, conjunctions are *italicized*. Past tense, perfect aspect, and progressive verbs are **bolded**, relative clauses with *that* are in [square brackets], *that*-clauses controlled by non-factive/communication verbs and non-finite *to*-clauses controlled by verbs of modality/ causation/effort and desire are underlined:

- 7.10 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. [POLISCI-QL]
- 7.11 At the beginning of this interaction, Student 1 **made** a mistake with the adjective ending to groB, a grammatical feature₁ [that **had not yet been covered** in the class.]₁ Furthermore, Student 1 **did not use** the preferred German adjective for long. Florian **provided** an implicit lexical recast, using the preferred German adjective for this meaning, *and* simultaneously **supplied** the correct adjective ending. He later **noted**, however, that this was not an intentional corrective recast. [AL-QL]

In these two excerpts, the focus is on describing, in detail, an account of events, and these events serve as the evidence on which an analysis is built. In the qualitative research paradigm, the purpose is to describe the natural course of events or actions, and then build interpretations upon those observations. Thus, there is a focus on establishing a narrative that sets up a reconstructed event to serve as evidence for the writers’ claims and interpretations. Past tense, perfect aspect, aspectual verbs, and time adjectives are particularly common in establishing the sequence and timing of events and happenings, while communication verbs are used to report speech from the often human actors that are represented by 3rd person pronouns, group nouns, and animate nouns.

The negative pole of this dimension, on the other hand, is characterized by the use of technical, quantity, and concrete nouns, as well as passive voice verbs. Theoretical and quantitative physics have the largest negative dimension scores for Dimension 3, and excerpts 7.12 and 7.13 show how these features are employed

in theoretical physics. Technical, quantity, and concrete nouns are **bolded**, passive main verbs are underlined, and attributive size adjectives are *italicized*:

- 7.12 Therefore, S_z is a conserved **quantity**. As a result, in terms of the standard basis vectors [symbols] and [symbol], the reduced two-spin density matrix, which is constructed by tracing out the rest spin degrees of freedom, can be written as [formula]. In terms of the spin correlation functions, the elements of $i, i+1$ in **equation** (2) are given by [formula]. Obviously, by its definition, [formula] is a semi-positive definite matrix. Let [formula] be its eigenvalues. We define the two-site local entanglement E_v of the system to be the 4 von Neumann entropy of [formula], i.e., we have [formula]. As is well known, on a d -dimensional simple cubic lattice, the phase **diagram** of the antiferromagnetic XXZ model is divided into three parts by two phase transition points $1 = -1$ and $2 = 1$. [PHYS-TH]
- 7.13 In writing [equation], the second term, F is of *lower* order in powers of n compared to the first one, since D is of order n . (Recall that $D^2 = -Cn$ is of order n .) Thus N is the leading term we need at *large* n , the corrections to (57) are subdominant. They will be needed for *higher* dimensional spaces, as we shall see later. [PHYS-TH]

Excerpt 7.12 illustrates one of the key characteristics of theoretical physics with respect to Dimension 2: the presentation of evidence using mathematical formulas. This focus on mathematics clearly relates to the the nouns and adjectives that seek to quantify or describe the size of objects (which also load on this dimension). In addition, the passive voice is used frequently in the prose introducing and explaining the formulas, and functions to establish the procedural steps in the analysis related to the computations.

Quantitative physics has a Dimension 2 score as low as theoretical physics, and while some evidence is presented as mathematical formulas, this use is not as prevalent in quantitative physics. Excerpt 7.14 illustrates that in addition to passive voice being used to establish steps in the analysis or procedural methodology, the nature of the discipline of physics is also explanatory, as the discipline is focused on presenting quantitative displays of evidence in figures, tables, and some formulas.

- 7.14 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. These events are associated with the double escape **peak** of the 2615 keV 208 Tl photon. The photon produces electronpositron pairs of which the positron subsequently annihilates. [PHYS-QT]

Quantitative biology also has a negative dimension score on Dimension 2, yet there is an almost 10 point difference between the physics registers. In biology, the primary focus within the context of this dimension is in the description of the methodological steps carried out by the researchers. Biology typically relies on quantitative displays of evidence, but does not use explications of mathematical formula in its arguments, one area in which physics relied on these negative features:

- 7.15 After selecting relevant variables using a stepwise procedure, a partial regression analysis was carried out only with those variables that were consistently selected by most richness and diversity indices (Table 2). 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. [BIO-QT]

Quantitative political science and applied linguistics have low positive dimension scores, and an examination of research articles in these two categories reveals that the quantitative social sciences rely on both the positive and negative features of Dimension 2. This may be the result of two competing factors. First, there could be some influence of discipline, since the qualitative registers in these two disciplines have the highest dimension scores on Dimension 2. Second, as quantitative research, these two registers also share characteristics related to research paradigm with the quantitative sciences. That is, these quantitative political science and applied linguistics still have an underlying focus on human subjects or events (which inherently involve people), and they exhibit tendencies to provide elaborated introductions and literature reviews in which they establish the scene for research (often through narration of a situation) and place themselves in the existing body of literature. This is illustrated in excerpt 7.16, where 3rd person pronouns, group nouns, and animate nouns are underlined, past tense, perfect aspect, and progressive verbs are **bolded**. Relative clauses with *that* are in [square brackets], *that*-clauses controlled by non-factive/communication verbs and non-finite *to*-clauses controlled by verbs of modality/ causation/effort and desire are underlined. Passive voice main verbs are underlined and bolded:

- 7.16 *Quantitative Political Science (Ramírez 2007):*
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 (Fullerton, 1997; Fullerton & Toossi, 2001; Passel & Suro, 2003; Vernez, 1998, 1999; Vernez, Krop, & Rydell, 1999). Less has

been done to understand the contemporary and future ramification of the influx of both native-born and naturalized Latinos in the electorate or their respective patterns of participation. In 1980, there **were** 2.5 million Latinos voters in the United States. By 2000, this figure **had more than doubled** to 5.5 million voters. Although Latinos **composed** less than 6% of the electorate in 2000,¹ the growth rate of that group **has been** impressive.

However, quantitative political science and applied linguistics also carry the methodological values of quantitative research to exercise control or research contexts and to produce concise, informative descriptions of the procedures followed to collect and analyze data. There are many instances of passive voice verbs (negative Dimension 2 features) to report research procedures, as well as past tense and perfect aspect verbs (positive Dimension 2 features) to report the results of studies:

7.17 *Quantitative Applied Linguistics* (Siyanova & Schmitt 2008):

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. Although we **were** interested in judgements about the acceptability of the collocations, collocations₁ [that are used frequently by natives]₁ are clearly acceptable, while collocations₂ [that do not occur in 100 million words]₂ are much less likely to be so. We felt that a judgement task relating to frequency would be more transparent to our participants than a task asking them to rate acceptability. Therefore, the instructions **required** the participants to rate the collocations according to frequency on a six-point scale.

In the next section, we see yet another pattern with respect to the organization of disciplines and registers along a dimension of variation. Dimension 3 cuts across disciplines which have at their foundation an inquiry into the workings of the human mind and disciplines with other areas of inquiry.

7.4.3 Dimension 3: Human vs. non-human focus

Dimension 3 consists of 11 positive features and 4 negative features. The positive features include 2nd and 3rd person pronouns, process nouns, mental, activity, and communication verbs, and progressive aspect. In addition to general wh-clauses, several stance features also characterize the positive end of this dimension: *that*-clauses controlled by factive verbs, and *to*-clauses controlled by verbs of desire and speech verbs. There are only four negative features on this dimension: all attributive adjectives, attributive adjectives indicating topic, general adverbs, and prepositions.

Figure 7.4 shows the distribution of the disciplines and registers along Dimension 3. Dimension 3 appears indicate a dichotomy between two types of disciplines: (1) disciplines which have human beings and their mental/cognitive activities at the

heart of their subject domain (applied linguistics and philosophy), and (2) disciplines whose topic domain is not focused on the cognitive activities of human beings. It should be noted that because there are few variables on the negative set of features for Dimension 3, interpretations of this part of the dimension is limited and preliminary.

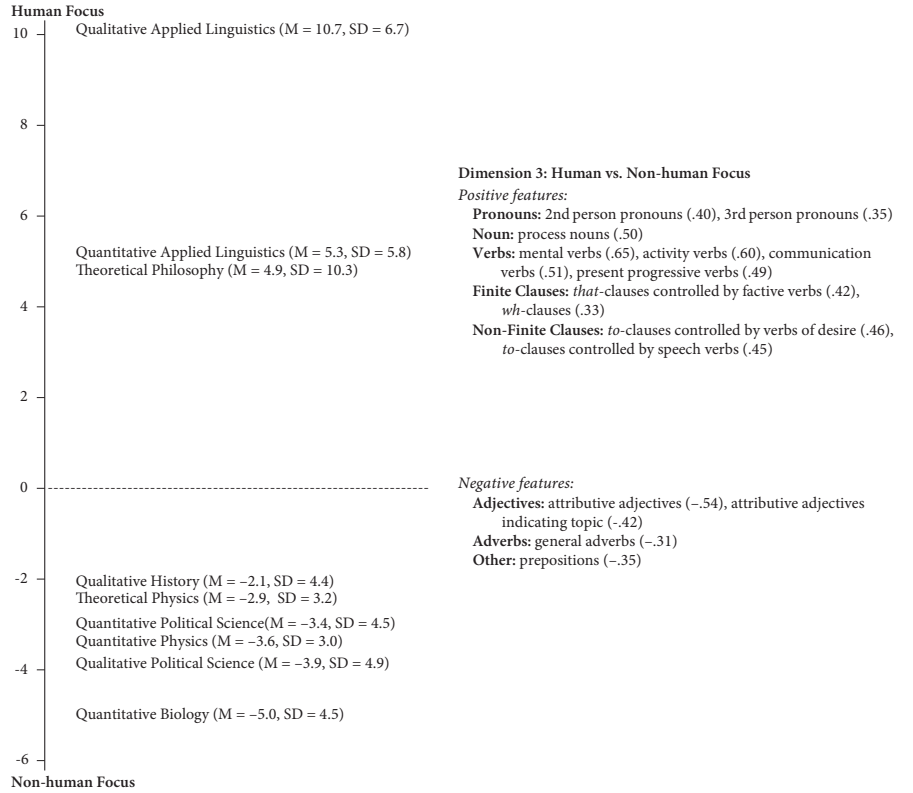


Figure 7.4. Distribution of disciplines and registers along Dimension 3: Human versus non-human focus. One-way ANOVA results: $F = 28.42$, $p = .000$, $r^2 = .47$. Post-hoc comparisons are listed in Appendix F

As the situational analysis in Chapter 4 summarized, the discipline of applied linguistics is concerned with humans and language, with the ways in which we use language, how we learn and acquire language, and the ways of teaching language to promote language acquisition (to name a few). Philosophy, on the other hand, is not focused on language specifically, but on the broader human mind and how we can use logic to understand human nature, human problems, and human cognition.

Reflecting the nature of the object of inquiry in these two disciplines, Dimension 3 shows parallels between the use of a variety of verbal structures that quite

often take human agents as the subjects of the verbs. These verbal structures include semantic categories of verbs that we use to describe mental processes, activities, and communicative acts. They also include stance clauses, and the human agents can be explicitly used as the subject (verbs marked 2, 4, 5, 6), or implied based context (verbs marked 1, 3, 7):

- 7.18 Teachers built background knowledge by **asking**_[1 – implied ‘teachers’] learners what they **thought**_[2 – explicit ‘learners’] the site would be about [AL-QL]
- 7.19 Indeed, immunology **is often described**_[3 – implied ‘people’] as the science of discrimination between self and non-self [PHIL-TH]
- 7.20 We **discuss**_[4 – explicit authors as ‘we’] these patterns in turn below [AL-QL]
- 7.21 Egalitarians can be pluralists about value. They **think**_[5 – explicit egalitarians as ‘they’] that there is a reason to level down—where doing so will make a distribution fairer—but **recognize**_[6 – 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**_[7 – implied ‘we’ as humankind] people [to be better rather than worse off.] [PHIL-TH]

In addition, the examples illustrate the variety of roles that these explicit and implied human agents can have. For example, mental, activity, and communication verbs, as well as *to*-clauses controlled by desire and speech verbs, are used to portray the thoughts and ideas of participants in the research (excerpts 7.22–7.24). In philosophy, this is often in the form of unreal characters (as in excerpt 7.25), or indefinite pronouns (excerpt 7.26) to refer generally to human beings that are used in the article to explore and illustrate logical processes.

- 7.22 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. [AL-QL]
- 7.23 After **reading** these instructions, each participant completed a series of practice sentences (see Appendix D) and **made** practice grammaticality judgements for these sentences. [AL-QT]
- 7.24 Henry **discussed** how good writers from another elite university in the same city mediated his writing although he **knew** none of them [AL-QL]
- 7.25 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. [PHIL-TH]
- 7.26 Memory is ordinarily taken to be factive. One cannot **remember** that which did not happen. [PHIL-TH]

While these examples have subjects referring to the entities being studied, these same verbs and structures are also often used to make connections to previously established theories and findings, as in excerpts 7.27 and 7.28. This second use is not unique to philosophy and applied linguistics, although a preliminary analysis suggests that this use is highly prevalent in these disciplines as the writers provide extensive theoretical grounding for both introducing studies and concepts, as well as interpreting results.

- 7.27 Strawson **has argued** that our ordinary conception of moral responsibility requires a kind of ultimate self-creation that is incoherent. Strawson **gives** various different formulations of the argument, but I **find** the versions presented in his article, “The Bounds of Freedom,” particularly lucid. [PHIL-TH]
- 7.28 Goss, Ying-Hua and Lantolf (1994), who **compared** grammatical judgement tasks **completed** individually and in pairs by learners of Spanish, **found** modest differences in favour of pairs and only on some grammatical features. [AL-QT]

Finally, these same structures are used with the authors/researchers as the grammatical subjects of the verbs. Again, this use is not limited to philosophy and applied linguistics by any means. However, the situational analysis presented in Chapter 4 indicated that these two disciplines are among those that most explicitly and extensively state the purpose of the research, as well as discuss the nature of data and procedures (particularly applied linguistics).

- 7.29 In order to **see** whether applying the cognitive typology can be of assistance in resolving some problems in a particular context of translation, I **analysed** original transcripts of police interviews with Spanish-speaking witnesses and suspects, with translation into English **provided** by certified court interpreters. [AL-QL]
- 7.30 I **argue** for an alternative justification for conservation in the capacity of foresight, which **requires** us [to act not only upon duties that we have now, but also upon those that we will predictably have in the future.] [PHIL-TH]
- 7.31 We **noticed** two aspects that were not **addressed** in the literature. First, none of the prereading methods **studied explored** the possibilities of content-area materials available from authentic texts within the discipline. While EAP reading at the higher level may be more of a reading problem than a language problem, we **believe** it is worthwhile to **explore** the utility of content materials for EAP reading intervention. [AL-QT]

As noted in the discussion above, not all of these uses of mental, activity, and communication verbs are limited to use in philosophy and applied linguistics. In addition, we could also argue that history and political science, in studying historical

events and social organization are also inherently human-based disciplines. We might assume that because the events, processes and situations that are described in history and political science are carried out by human beings, these disciplines would look more similar to applied linguistics and philosophy than is demonstrated by Figure 7.5. Yet, history and political science are generally *not* significantly different (see Appendix F for post-hoc analyses) from the hard sciences along this Dimension. Therefore, it seems that there is a fundamental difference between the two disciplines that explore mental processes or phenomena that are connected to the cognitive abilities of humans and other disciplines. History and political science fall in with physics and biology along this dimension because object of study is typically not on humans and their cognitive processes or communicative roles, but rather on the *events* and *situations* that make up human history and society.

To illustrate, we can look at the following excerpts, one from a history article and one from a quantitative political science article. These two excerpts illustrate that although the phenomena under investigation are events and trends carried out by human beings, the focus of these analyses is not on understanding the cognitive characteristics of human beings, but rather practices, actions/events, and so on:

7.32 *Qualitative History (Fette 2007):*

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. Women were in fact only one unwanted social category among many in the French professions; foreigners, naturalized citizens, and the lower social classes also served as scapegoats for supposed overcrowding and loss of tradition. Of course not all women in pursuit of medical or legal careers were bourgeois and French, and thus faced multiple prejudices.

7.33 *Quantitative Political Science (Pacek & Radcliff 2008):*

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. Citizens are “emancipated” from the market in the sense that they can freely opt out of work, when necessary, without risking their jobs, incomes, or general welfare. [POLISCI-QT]

In sum, Dimension 3 distinguishes between disciplines which have humans, their thoughts, knowledge, and communication practices at the heart of their evidentiary practices, and disciplines which focus on either historical events and

conditions (history and political science) or the non-human natural world (biology and physics).

7.4.4 Dimension 4: ‘Academese’

Dimension 4 is characterized by the use of the fewest features, with 8 positive features and only 1 negative feature. The positive features include three sets of abstract nouns: process nouns, other abstract nouns, and nominalizations. Along with these three types of abstract nouns, existence verbs and adjectival structures (relational adjectives, *that*-clauses controlled by likelihood adjectives, and *to*-clauses controlled by all stance adjectives) also load positively on Dimension 4. Because there are so few features on this dimension, the interpretations offered here will be brief and are preliminary in nature. In fact, it is even more important to characterize the functional interpretation of this dimension while also considering the distribution of registers along the dimension. Figure 7.5 displays the mean scores for each discipline and register for Dimension 3.

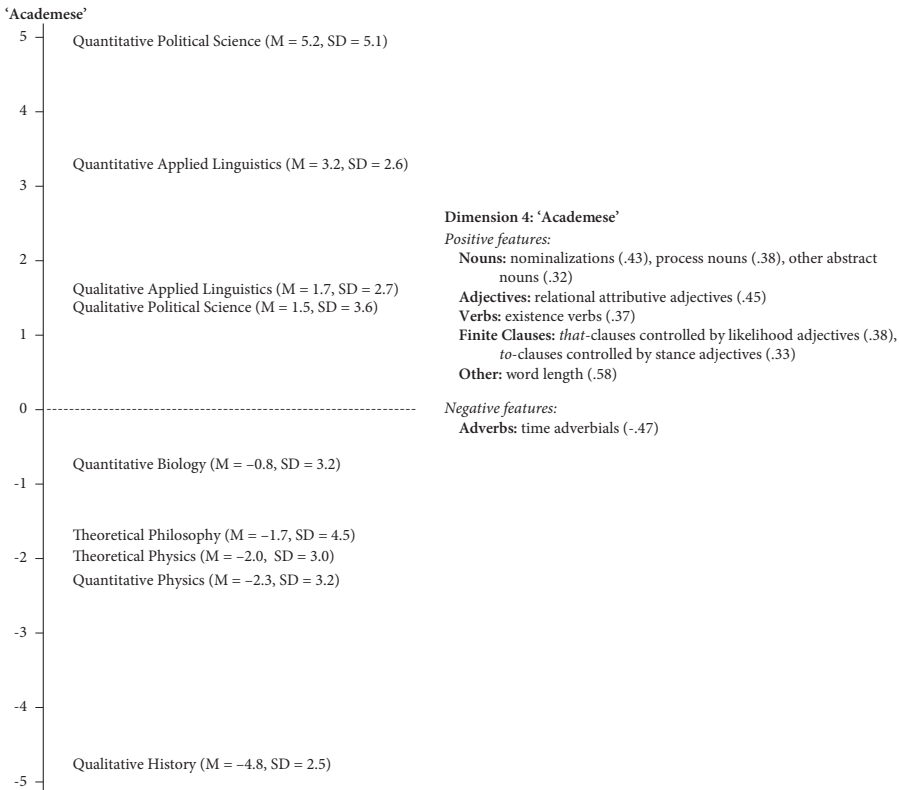


Figure 7.5. Distribution of disciplines and registers along Dimension 4: ‘Academese’. One-way ANOVA results: $F = 24.33$, $p = .000$, $r^2 = .43$. Post-hoc comparisons are listed in Appendix F

Along this dimension, only four registers have positive scores: quantitative political science and applied linguistics have the highest positive dimension scores, followed by qualitative political science and applied linguistics. Qualitative history has the highest negative dimension score, followed by quantitative and theoretical physics and theoretical philosophy. When we compare the situational characteristics of these two groups of disciplines (those with positive scores and those with negative scores), we see that the major situational difference between the two groups is the extent to which the disciplines are concerned with overtly representing themselves as empirical research through the inclusion of situational features like explicitly stated research designs, research questions, integrated citations, and labeling data and research processes.

To take a step back, the two theoretical disciplines (philosophy and physics) are clearly not empirical research paradigms, as evidenced by the very definition of 'theoretical' texts. However, during the interview processes described in Chapter 2, it came out that academics within the disciplines of history and quantitative physics, although technically empirical in nature (as they observe data which they use to draw conclusions from), would not necessarily overtly use the term 'empirical' to describe research practices in the disciplines (T. Porter, personal communication, May 6, 2009; G. Lubick, personal communication, January 26, 2010).

Thus, since the negative pole of this dimension is populated by the two theoretical registers, along with qualitative history and quantitative physics, it is possible that this dimension is distinguishing between disciplines and registers which are explicitly promoted as 'empirical' in nature and those which are less concerned with being labeled empirical (regardless of whether or not they are empirical in nature, as is the case for qualitative history and quantitative physics). It appears that registers which are expressly concerned with packaging themselves as 'scientific' inquiry (such as the social sciences) employ features such as abstract nouns, process nouns, and nominalizations, mimicking the process of grammatical metaphor that Halliday's work has documented so extensively in scientific discourse.

If this is the case, it is interesting to note that the two quantitative social sciences (applied linguistics and political science) have the highest positive scores on this dimension, followed by the two qualitative registers in these same disciplines. When we look at the features that characterize the positive end of this dimension of variation, we see several adjectival structures, such as relational adjectives (e.g., *basic, common, different, general, individual, main, particular, same, similar, various*), existence verbs (e.g., *appear, contain, defined, exist, illustrate, include, indicate, reflect, represent, stay, tend, vary*). In addition, *to* and *that*-clauses headed by certain stance adjectives are also important for this dimension, as well as nominalizations and process and other abstract nouns. These patterns

are illustrated in the two excerpts below (positive features **bolded**), and we can see that these existence verbs and adjectival structures are often used to describe and interpret the results of the data analysis, a discourse function highly associated with empirical research.

- 7.30 All round then, given the **various** terms and working **definitions** used in **studies** of recasts, it may be **unhelpful** to lump together all the **different** types under the single label of 'recasts' and also to assume that all recasts are categorically implicit. [AL-QT]
- 7.31 On the issue of **same sex marriage**, uninformed women were approximately 20 per cent more **likely** to support **same sex marriage** than were men, and **information** had the effect of widening that gap, bringing the probability of **support** up to nearly 29 per cent **higher** than men. **Information** had a **similar** (and **larger**) **effect** on the **support** for easier access to **abortion**... the informed nonreligious are 28 per cent more **likely** to support **same sex marriage** than are the informed religious and 34 per cent more **likely** to support **abortion**. [POLISCI-QT]

Thus, although this analysis is preliminary, it seems that dimension 4 distinguishes between disciplines and registers that are overtly empirical in nature, and the linguistic features on this dimension help researchers characterize and make sense of results in order to offer interpretations of that data.

7.5 Conclusions

The new multi-dimensional analysis carried out in this chapter has revealed that linguistic variation in published academic research articles varies according to multiple parameters. 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 unrecognized or disregarded in previous studies of disciplinary variation. Parameters such as the nature of evidence, the presence or absence of data, research methods (qualitative vs. quantitative vs. theoretical research), and the object of study. That is, 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. Each of these characteristics corresponds to specific linguistic patterns, and linguistic resources work together to construct meaningful discourse that is reflective of the nature of the discipline as well as the research paradigm within which the scholarly work fits.

In addition, the distinct ways in which the registers and disciplines distributed along the four dimensions highlights the complex nature of academic writing.

Writers utilize a multitude of linguistic resources that correspond to an equally complex set of situational, or non-linguistic, characteristics. For example, articles in theoretical physics both 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, as illustrated in Dimension 1. Likewise, quantitative applied linguistics articles utilize both moderately dense expression of information like scientific writing more generally (Dimension 1), a certain degree of narrative orientation (Dimension 2), along with a major focus on human actors (Dimension 3).

Somewhat surprisingly, it turns out that one of those dimensions of variation continues to highlight the differing degree to which texts balance between 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. Despite a few differences in the specific features that loaded on Dimension 1 in the present analysis, the degree of correspondence also supports the idea that this dimension of variation is a universal construct in describing register variation in English. More specifically, this new dimension reflects those grammatical complexity features which are most important for professional academic writing specifically.

The MD approach is unique because the underlying patterns of variation are derived *inductively* and *quantitatively* from the analysis of the corpus, and then interpreted functionally to explain the observed patterns. In addition, the dimensions of variation that emerge from the analysis are typically considered previously unrecognized constructs (see Biber 2010). In this approach, no organization of the texts is overtly placed on the corpus during the statistical process of identifying co-occurrence patterns. Such organization of texts into sub-corpora to calculate per-register/discipline dimension scores is carried out only after the co-occurrence patterns have been established based on the inductive analysis of the corpus.

In the next chapter, I consider the results of the three analyses presented in Chapters 5, 6, and 7, with the goal of synthesizing the results into statements about what we know about these disciplines, and linking these findings more explicitly to the situational characteristics of the disciplines and registers that were described in Chapter 4.