CHAPTER 7

Multi-dimensional patterns of variation among university registers

7.1 Introduction

The preceding chapters have described the functions of individual linguistic features in university registers. These descriptions identify many important linguistic differences across registers, and they enable detailed consideration of the functions served by particular features. Taken together, these analyses suggest several general patterns of variation among university registers, including:

- spoken registers are systematically different from written registers, with respect to a wide range of vocabulary characteristics and lexico-grammatical features
- interactive registers are different from monologic registers
- academic registers are different from the behavior-management (directive) registers, whether spoken or written

The present chapter shifts to a wider perspective, using Multi-Dimensional (MD) analysis to study the overall patterns of register variation among university texts. In MD analysis, the distribution of a large set of linguistic features is analyzed in a multi-register corpus of texts. Specifically, factor analysis is used to identify the systematic co-occurrence patterns among these linguistic features – the 'dimensions' – and then texts and registers can be compared along each dimension.

Multi-Dimensional analysis provides a complementary linguistic perspective on university registers to the one developed in preceding chapters. That is, earlier chapters in this book have focused on particular linguistic levels or domains, considering linguistic features belonging to a single structural level (e.g., word choice; part-of-speech categories; dependent clause types), or considering features that all serve the same general function (e.g., expressing stance). These chapters have employed a comparative approach and quantitative analysis to determine the relative distribution of individual linguistic features. By using quantitative comparisons to the range of other university registers, the descriptions are able to determine whether a given frequency of occurrence is notably common or rare in a target register. This quantitative comparative approach treats register as a continuous

construct: texts are situated within a continuous space of linguistic variation, enabling analysis of the ways in which registers are more or less different with respect to the full range of core linguistic features.

It turns out, though, that the relative distribution of common linguistic features, considered individually, cannot reliably distinguish among registers. There are simply too many different linguistic characteristics to consider, and individual features often have idiosyncratic distributions. That is, although the distributions of individual features are interpretable in functional terms, these individual patterns cannot be used to determine the extent to which any two registers are similar or different. Rather, overall register differences are describable with respect to sets of co-occurring linguistic features.

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

Although this general theoretical perspective has been widely accepted, linguists lacked the methodological tools required for such analyses before the availability of corpus-based techniques. The MD approach was developed to analyze the linguistic co-occurrence patterns associated with register variation in empirical/quantitative terms. Early MD studies investigated the patterns of variation among general spoken and written registers in English (Biber 1985, 1986, 1988), while subsequent studies documented the patterns of register variation in other languages (see, e.g., Biber 1995), or in more specialized discourse domains (see, e.g., Conrad & Biber 2001).

MD analysis uses the power of multivariate statistical techniques to investigate the quantitative distribution of linguistic features across texts and registers. Linguistic co-occurrence is analyzed in terms of underlying 'dimensions' of variation which are identified quantitatively, by a statistical factor analysis, rather than on an a priori functional basis. The dimensions resulting from MD analysis have both linguistic and functional content. The linguistic content of a dimension is a group of features (such as nouns, attributive adjectives, prepositional phrases) that co-occur regularly in texts. On the assumption that co-occurrence reflects shared functions, these co-occurrence patterns are interpreted to assess the situational, social, and cognitive functions most widely shared by the linguistic features.

7.2 Application of the 1988 MD analysis to university registers

7.2.1 Overview of the 1988 MD analysis

The first major MD analysis (Biber 1988; see also Biber 1985, 1986) was undertaken to investigate the relationship among general spoken and written registers in English. The corpus included 481 spoken and written texts of contemporary British English, taken from the Lancaster-Oslo-Bergen Corpus and the London-Lund Corpus. These texts were sampled from 23 major register categories, including academic prose, press reportage, fiction, letters, conversations, interviews, radio broadcasts, and public speeches.

The 1988 study used factor analysis to identify the groups of linguistic features associated with each dimension; these are the sets of linguistic features that co-occur in texts with markedly high frequencies. The factors are interpreted as functional 'dimensions' based on the assumption that linguistic features co-occur in texts because they reflect underlying shared communicative functions. 67 linguistic features were analyzed (e.g., first and second person pronouns, nominalizations, past tense verbs, *that* relative clauses, etc.), and a principal factor analysis was used with a 7-factor solution to identify the 'dimensions' of variation. (Details of the analysis are provided in Biber 1988: Chapters 4 and 5; 1995: Chapter 5.)

The first five factors from the 1988 factor analysis are readily interpretable as 'dimensions' of register variation, based on consideration of the linguistic features co-occurring on each dimension together with the similarities and differences among registers with respect to the group of features. Interpretive labels have been proposed for each dimension, reflecting the underlying functions:

1. Involved versus Informational Production
2. Narrative versus Non-narrative Concerns
3. Situation-Dependent versus Elaborated Reference
4. Overt Expression of Persuasion
5. Non-Impersonal versus Impersonal / Abstract Style

7.2.2 Relations among university registers with respect to the 1988 dimensions

Although these dimensions were identified from analysis of a general corpus of spoken and written texts, they can be applied to the description of more restricted discourse domains, including university spoken and written registers. Such an analysis is based on the premise that the dimensions identified in the 1988 study represent general linguistic/functional parameters of variation applicable to more specialized discourse domains. That is, because the 1988 analysis was based on a broad sample of texts and registers, and included a large sample

of linguistic features from several structural and functional levels, the dimensions have been regarded as valid for the description of register variation in more specialized domains.

Biber, Conrad, Reppen, Byrd, and Helt (2002) apply the 1988 dimensions to the analysis of the university spoken and written registers in the T2K-SWAL Corpus. The results of that study show that students must negotiate registers representing a tremendous range of linguistic variation. On all dimensions except Dimension 2 (narrative discourse), university registers fell at both ends of these linguistic parameters. Students must deal not only with informationally dense prose but also with interactive and involved spoken registers. They must handle texts with elaborated reference as well as those that rely on situated reference, and texts with features of overt persuasion as well as texts that lack those features. They must understand discourse that uses an impersonal style (with many passives) as well as discourse that tends to avoid passives.

The distribution of registers along Dimension 1 – Involved versus Informational Production – is especially noteworthy. Academic registers are typically assumed to be extremely informational, but it turns out that university students also encounter academic registers that are highly interactive and involved. Even registers with a strongly informational purpose, such as classroom teaching and study groups, are marked for the features of face-to-face interaction and a focus on personal stance, rather than the features of informational production.

The 2002 MD study found that most of these dimensions show a strong polarization between spoken and written registers. The written registers – regardless of their specific purpose – are characterized by informationally dense prose (Dimension 1), a non-narrative focus (Dimension 2), elaborated reference (Dimension 3), few features of overt persuasion (Dimension 4), and an impersonal style (Dimension 5). (The only exception to this pattern is the course management register, which uses the Dimension 4 features of overt argumentation.) In contrast, the spoken registers – again regardless of purpose – are all characterized by features of involvement and interaction, situated reference, more overt persuasion, and fewer features of impersonal style. No register in the T2K-SWAL Corpus is characterized by a narrative focus (Dimension 2).

This general pattern of polarization contrasts with previous MD studies of English, which did not find spoken and written registers to be so sharply distinguished. For example, fiction writing is similar to many spoken registers (see Biber 1988: Chapter 7): fiction uses the features of both involved and informational production (Dimension 1), and it is marked strongly for the use of narrative features (Dimension 2), situation-dependent reference (Dimension 3), and non-impersonal style (Dimension 5). In contrast, no written register in the T2K-SWAL Corpus was similar to any of the spoken registers with respect to the 1988 dimensions.

The sharp divide between spoken and written university registers is especially surprising given the wide range of communicative purposes represented by the registers in the T2K-SWAL Corpus. The spoken registers, for example, range from interpersonal interactions with both social and informational purposes (e.g., service encounters and study groups), to monologic discourse with a primary informational focus (e.g., classroom teaching). However, despite these differences, all spoken university registers are sharply distinguished from written university registers with respect to the 1988 dimensions of variation.

These research findings raise the question of whether other dimensions of variation might reflect differences in purpose or task, and so distinguish among academic versus non-academic registers within the spoken and written modes. That is, the description of university register variation in Biber et al. (2002) is based on the dimensions identified previously in the 1988 MD study; those dimensions were derived from analysis of a general corpus of spoken and written registers (e.g., conversation, interviews, newspaper reportage, editorials, fiction, and academic prose). It might be the case, however, that additional dimensions of variation are important for distinguishing among university spoken and written registers. To investigate this possibility, a new factor analysis was carried out to identify the dimensions of variation that actually occur in the T2K-SWAL corpus.

7.3 Motivating a 'new' MD analysis of university registers[1]

The application of the 1988 dimensions to new discourse domains has been highly productive, enabling descriptions of the similarities and differences among specialized registers within that multi-dimensional linguistic space. This approach has been used in several previous studies of specialized registers (see Conrad & Biber 2001, and the survey of studies in Biber to appear); these include direct mail letters (Connor & Upton 2003), non-profit grant proposals (Connor & Upton 2004), author styles (Connor-Linton 1988, 2001; Biber & Finegan 1994b), conversational registers (Helt 2001; Quaglio 2004), and the speech of women and men in dramatic dialogue (Rey 2001; Biber & Burges 2000). As summarized in the preceding section, this same approach has also been used to investigate the overall patterns of variation among university registers.

However, this approach does not enable description of the dimensions that are actually most important in a particular domain of use. That is, linguistic features can co-occur in particular ways in different discourse domains, reflecting the specialized functional priorities of those domains.

An alternative approach is to carry out a completely new MD analysis: a new factor analysis to identify the co-occurrence patterns that actually occur in a corpus of texts. This approach is appropriate when analyzing a new discourse

domain that includes many different text categories. Conducting a 'new' multi-dimensional analysis allows identification of the co-occurrence patterns specific to a given domain, and registers can then be compared with respect to those 'new' dimensions. The present chapter adopts this approach to identify the underlying dimensions of variation that distinguish among university registers, disciplines, and levels.[2]

The Multi-Dimensional analysis presented here is further different from the 1988 analysis in that it incorporates a much larger set of linguistic features, building on the detailed descriptions of linguistic features presented in Chapters 4–6. The study began with 129 linguistic features (see Appendix A), although these were subsequently reduced to 90 features that were conceptually distinct and functioned in statistically meaningful co-occurrence patterns. The following section (7.4) presents the statistical factor analysis used for this description, and then in Section 7.5, I interpret these dimensions to describe the patterns of variation among university registers.

7.4 Factor analysis of linguistic features in the T2K-SWAL corpus

The factor analysis of the T2K-SWAL Corpus was based on the full set of linguistic features introduced in Chapter 2 and Appendix A. Methodologically, the procedure follows the steps outlined in previous studies (Biber 1988: Chapters 5–6; Biber 1995: Chapter 5; Conrad & Biber 2001: Chapter 2). As described above, this analysis differs from most MD studies in that it is based on a separate factor analysis for this corpus of texts, rather than applying the 1988 general model of variation.

Only 90 of the original 129 linguistic features were retained in the factor analysis. Some features were dropped because they overlapped to a large extent with other features. For example, the original 129 features included counts for 'high frequency verbs', 'high frequency nouns', and 'high frequency adjectives'; and they also included counts for semantic classes of nouns, verbs, and adjectives (e.g., 'mental verbs' or 'communication verbs'). However, it turned out that the specific words included in the high frequency classes overlapped extensively with the words included in the semantic categories. For example, verbs like *think* and *know* are among the most important high frequency verbs and also important members of the 'mental verb' category. Thus, to a large extent, the high frequency categories were measuring the same constructs as the semantic category distinctions, and they were therefore dropped from the factor analysis.

In other cases, features were dropped because they were extremely rare. For example, the original set of features included separate counts for each semantic subclass of phrasal verb (such as 'intransitive activity phrasal verb' and 'transitive

communication phrasal verb'). These subclasses were generally rare, and so they were all combined into a single feature: 'phrasal verbs'.

Finally, some features were dropped because they shared little variance with the overall factorial structure. ('Communality estimates' produced by the statistical analysis indicate the extent to which a given feature participates in the overall pool of shared variance accounted for by the factor analysis. In general, features with communalities below .15 do not have meaningful factor loadings on any factor.)

The solution for four factors was selected as optimal, and the factors were rotated using a Promax rotation. Solutions with fewer factors resulted in a collapsing of linguistic features onto single factors, making the interpretation of those factors difficult. Solutions with additional factors accounted for little additional variance and those factors were represented by only a few features. Biber (2003b: Appendix I) presents the full factorial structure of this analysis; the discussion here focuses on the interpretation of these dimensions.

Table 7.1 summarizes the important linguistic features defining each dimension, including only features with factor loadings larger than + or – .3. Consideration of these features provides the basis for the factor interpretations as 'dimensions' of variation, identifying the shared communicative functions underlying each set of co-occurring features.

These dimensions can be used to analyze the linguistic characteristics of texts by computing 'dimension scores' (or 'factor scores') for each text: a summation of the standardized frequencies for the features with salient loadings on a dimension. Only features with loadings greater than |.30| on a factor are used in the computation of dimension scores. For example, the Dimension 1 score for each text is computed by adding together the frequencies of contractions, demonstrative pronouns, pronoun *it*, first person pronouns, present tense verbs, etc. – the features with positive loadings on Factor 1 (from Table 7.1) – and then subtracting the frequencies of nominalizations, word length, moderately common nouns, prepositions, etc. – the features with negative loadings.

Before dimension scores are computed, the individual feature scores are standardized to a mean of 0.0 and a standard deviation of 1.0 (based on the overall mean and standard deviation of each feature in the T2K-SWAL Corpus). This process translates the scores for all features to scales representing standard deviation units. Thus, regardless of whether a feature is extremely rare or extremely common in absolute terms, a standard score of +1 represents one standard deviation unit above the mean score for the feature in question. That is, standardized scores measure whether a feature is common or rare in a text relative to the overall average occurrence of that feature. The raw frequencies are transformed to standard scores so that all features on a factor will have equivalent weights in the computation of dimension scores. If this process were not followed, extremely common features

Table 7.1 Summary of the factorial structure of the T2K-SWAL MD analysis (factor loadings are in parentheses)

Dimension 1: Oral vs. literate discourse

Features with positive loadings:
contractions (.91), pronouns: demonstrative (.91), pronouns: *it* (.87),
pronouns: 1st person (.81), verbs: present tense (.81), adverbials: time (.80),
adverbs: common (.80), pronouns: indefinite (.79), *that*-omission (.75),
discourse particles (.73), common verbs: mental (.73), lexical bundles: pronoun initial (.73),
stranded prepositions (.72), WH questions (.69), clause coordination (.69),
adverbial clauses: causative (.67), adverbials: place (.67), adverbs: moderately common (.66),
verbs: progressive (.65), WH clauses (.63), common verbs: activity (.58),
that-clauses: controlled by likelihood verbs (.58), lexical bundles: other (.58),
adverbials: certainty (.58), lexical bundles: WH word initial (.56), pro-verb *DO* (.56),
adverbials: hedges (.55), adverbial clauses: other (.55), pronouns: 2nd person (.53),
that-clauses: controlled by certainty verbs (.53), lexical bundles: verb initial (.50),
verbs: past tense (.50), adverbials: likelihood (.46), verbs: phrasal (.42),
common verbs: communication (.40), pronouns: 3rd person (.39),
lexical bundles: *it* initial (.35), *that*-clauses: controlled by communication verbs (.34),
adverbial clauses: conditional (.31)

Features with negative loadings:
nouns: nominalizations (–.95), word length (–.93), nouns: moderately common (–.90),
prepositional phrases (–.86), common nouns: abstract (–.82), adjectives: attributive (–.78),
passives: agentless (–.75), passives: postnominal (–.75), type/token ratio (–.67),
common adjectives: relational (–.67), relative clauses: WH; prep fronting (–.55),
passives: *by*-phrase (–.54), *to*-clauses: controlled by stance nouns (–.54),
phrasal coordination (–.50), relative clauses: WH; subject gaps (–.49),
common nouns: group (–.48), common adjectives: topical (–.44),
common nouns: human (–.37), common nouns: mental (–.36),
lexical bundles: preposition initial (–.36), common verbs: causative (–.34),
to-clauses: controlled by adjectives (–.33), adjectives: moderately common (–.30)

Dimension 2: Procedural vs. content-focused discourse

Features with positive loadings:
modals: necessity (.53), common verbs: causative (.51), verbs: moderately common (.48),
pronouns: 2nd person (.44), modals: future (.43), nouns: moderately common (.43),
to-clauses: controlled by verbs of desire (.42), common nouns: group (.41),
adverbial clauses: conditional (.35), *to*-clauses: controlled by other verbs (.35),
common verbs: activity (.33)

Features with negative loadings:
adjectives: rare (–.70), nouns: rare (–.63), adverbs: rare (–.49),
common verbs: simple occurrence (–.47), common adjectives: size (–.42),
verbs: rare (–.36), *to*-clauses: controlled by probability verbs (–.36),
passives: *by*-phrase (–.34), verbs: past tense (–.31), adverbs: moderately common (–.30)

Table 7.1 (*continued*)

Dimension 3: Reconstructed account of events
Features with positive loadings: pronouns: 3rd person (.63), common nouns: human (.60), *that*-clauses: controlled by communication verbs (.47), common verbs: communication (.45), verbs: past tense (.36), *that*-omission (.36), *that*-clauses: controlled by likelihood verbs (.36), common verbs: mental (.34), common nouns: mental (.32), *that*-clauses: controlled by stance nouns (.30)
Features with negative loadings: common nouns: concrete (–.53), common nouns: technical+concrete (–.49), common nouns: quantity (–.40)
Dimension 4: Teacher-centered stance
Features with positive loadings: relative clauses: *that* (.56), lexical bundles: preposition initial (.41), adverbials: certainty (.39), *that*-clauses: controlled by stance nouns (.36), adverbials: attitudinal (.36), adverbials: likelihood (.35), lexical bundles: noun initial (.35), adverbial clauses: other (.35), adverbial clauses: conditional (.30)
Features with negative loadings: WH questions (–.39), stranded prepositions (–.36)

would have a much greater influence than rare features on the dimension scores. The methodological steps followed to standardize frequency counts and compute dimension scores are described more fully in Biber (1988: 93–97).

Once a dimension score is computed for each text, the mean dimension score for each register can be computed. Plots of these mean dimension scores allow linguistic characterization of any given register, comparison of the relations between any two registers, and a fuller functional interpretation of the underlying dimension. Considering all five dimensions together enables comprehensive descriptions of the linguistic characteristics of particular registers and the linguistic similarities and differences among registers.

Table 7.1 (above) includes functional labels for each of the four dimensions:

Dimension 1: Oral vs. literate discourse
Dimension 2: Procedural vs. content-focused discourse
Dimension 3: Reconstructed account of events
Dimension 4: Teacher-centered stance

The following section presents the details of these functional interpretations, including a discussion of the co-occurring linguistic features grouped on each dimension, the distribution of university registers along each dimension, and a detailed consideration of the co-occurring features in particular texts. Then, in the concluding section, I compare the dimensions of variation found for the university

discourse domain to those found in the 1988 study for general spoken and written registers in English, discussing possible reasons for the observed similarities and differences.

7.5 Interpretation of the university (T2K-SWAL) dimensions of variation

7.5.1 Dimension 1: Oral vs. literate discourse

As in almost all previous MD analyses, the first dimension of the present analysis is associated with a fundamental oral/literate opposition. (Compare the surveys of other MD studies in Biber (1995, to appear).) Table 7.1 above shows that the positive features on Dimension 1 are associated with several major functional domains, including: interactiveness and personal involvement (e.g., 1st and 2nd person pronouns, WH questions), personal stance (e.g., mental verbs, *that*-clauses with likelihood verbs and factual verbs, factual adverbials, hedges), and structural reduction and formulaic language (e.g., contractions, *that*-omission, common vocabulary, lexical bundles). In contrast, the negative features are associated mostly with informational density and complex noun phrase structures (frequent nouns and nominalizations, prepositional phrases, adjectives, and relative clauses) together with passive constructions.

Figure 7.1 shows that all spoken registers in the T2K-SWAL Corpus have large positive scores on this dimension, while all written registers have large negative scores. At one level, this distribution is surprising given the major differences in purpose and planning across registers within each mode. That is, it might be expected that the informational-spoken registers – especially classroom teaching – might exploit the same styles of informational presentation as textbooks. However, with respect to Dimension 1 features, this is clearly not the case. Instead we see a fundamental opposition between the spoken and written modes here, regardless of purpose, interactiveness, or other pre-planning considerations. (Dimension 1 from the 1988 MD analysis is similarly associated with an absolute distinction between spoken and written university registers; see the discussion in Section 7.2.2 above.)

Service encounters, office hours, and study groups – the registers with the largest positive Dimension 1 scores – are all directly interactive; they are also the most conversational registers in the T2K-SWAL Corpus in terms of mixing involved, stance-focused personal purposes with the conveyance of topical information. Text Sample 7.1 illustrates the Dimension 1 characteristics of service encounters. Notice the dense use of 1st and 2nd person pronouns (*I, we, you*), contractions (e.g., *we're, don't, I'm, there's*), present tense verbs (e.g., *are, have,*

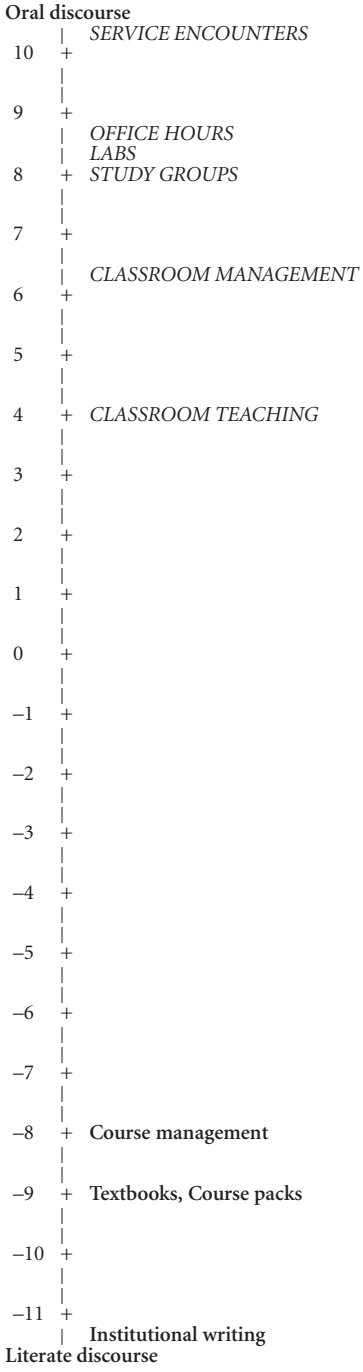

Figure 7.1 Mean scores of university registers along Dimension 1 – 'Oral vs. literate discourse'

get), time and place adverbials (e.g., *back, there, here, again*), indefinite pronouns (*something*), mental verbs (*think, want*), causative clauses, etc.

Text Sample 7.1: Service encounter; at the bookstore (servenbs_n125) [selected positive Dimension 1 features are in **bold underlined**]

Customer: Can **I** ask **you something**?
Clerk: Yeah.
Customer: **We're** at the previews and of course my book is **back there** with my husband. Do **you** have coupons?
Clerk: No **we** don't have any of them **here**. **You** guys only get them. Yeah.
Customer: OK.
Clerk: Did **you want** to come **back**? **Cos I** can hold onto your stuff.
Customer: Could **you** hold all this stuff? **Cos I know** if **I'm** getting a big sweatshirt there's one for a sweatshirt and one for a T. shirt.
Clerk: Yeah. **I'll** just hold onto them.
Customer: OK.
Clerk: **I'll** go ahead and just put them in a bag.
Customer: And **then we'll** just.
Clerk: Yeah **we're** open until six thirty. So.
Customer: **I think we're** gonna be **back there** registering at four or **something** and **then we'll** just come **back again**.

Office hours can be similarly interactive and involved. The following text sample illustrates many of the same positive Dimension 1 features as the service encounter (e.g., personal pronouns, present tense verbs, contractions, mental verbs). This sample also illustrates the use of discourse particles (*ok, well, oh*), progressive verbs (e.g., *was planning*), WH-questions (*what's up*), and WH-clauses (*what you did*):

Text Sample 7.2: Office hours (socpooh_n150) [selected positive Dimension 1 features are in **bold underlined**]

Instructor: let **me see** [unclear]
Student: **I need** help
Instructor: **you need** help **OK well then** let's talk
Student: yeah [laughing]
Instructor: **what's up**
Student: **well I I was planning** on just trying to do an outline of **what you did**
Instructor: {right, right
Student: and get the main points of realism
Instructor: mhm
Student: **OK I I** got it, **you know**, the several points
Instructor: **OK** and **we're** gonna talk about it more **today** if **you've** got some
Student: **oh**
Instructor: still some problems

Student: **well I I I well** maybe try to work on a thesis
Instructor: **right OK** let's **see what you did**

At the other extreme, institutional writing has the largest negative score on Dimension 1, making it even more 'literate' than textbooks or course packs. In fact, the linguistic style found in many university catalogs and program brochures is often more reminiscent of highly technical academic prose than textbooks written for novices in an academic discipline. The following program description for an anthropology major (Text Sample 7.3) begins with a friendly, inviting sentence having an extremely simple syntactic clause structure. However, this short sentence is immediately followed by complex sentences with multiple levels of clausal and phrasal embedding. Note especially the dense use of noun phrase structures, often with adjectives and prepositional phrases as modifiers.

Text Sample 7.3: Institutional writing (Web Catalog Academic Program Descriptions: Anthropology; otcatc.ant) [nouns, adjectives, and prepositions are in **bold underlined**]

PROGRAM DESCRIPTION.

Anthropology is the **study of people**. Its **perspective** is **biological**, **social** and **comparative**, encompassing all **aspects** of **human existence**, **from** the **most ancient societies to** those **of** the **present day**. **Anthropology** seeks to order and explain **similarities** and **differences between peoples of** the **world from** the **combined vantage points of culture** and **biology**.

Cultural and **Social Anthropology** deal **with** the **many aspects of** the **social lives** of **people around** the **world**, including our own **society**: their **economic systems**, **legal practices**, **kinship**, **religions**, **medical practices**, **folklore**, **arts** and **political systems**, **as well as** the **interrelationship of** these **systems in environmental adaptation** and **social change**. **Physical Anthropology** describes and compares **world human biology**. Its **focus** is **on humans** and the **primate order to** which they belong **as part of nature**, and it seeks to document and understand the **interplay of culture** and **biology in** the **course of human evolution** and **adaptation**. **Anthropological Linguistics** deals **with varied aspects of human language**, and the **characteristics of non human communication systems**, in order to achieve an **understanding of past** and **present human language systems** and their **significance in social life**.

Many of the negative features on Dimension 1 reflect the dense use of nouns and nominal modifiers in written informational texts; these features often occur together to build very complex noun phrase structures. For example, the second paragraph above begins with a very long sentence, which has only one main verb: *deal with*. Most of the sentence comprises a single noun phrase, functioning as the direct object of *deal with*. The sentence is marked up below to illustrate the extremely complex syntactic structure with multiple levels of embedding; head

nouns of noun phrases are underlined; the main verb is in **bold**; and brackets are used to delimit postnominal modifiers.

> Cultural and Social Anthropology **deal with** the many aspects [of the social lives [of people [around the world]]], [including our own society: [their economic systems, legal practices, kinship, religions, medical practices, folklore, arts and political systems], as well as [the interrelationship [of these systems [in environmental adaptation and social change]]]].

Textbooks are similar to institutional writing in their reliance on these 'literate' Dimension 1 features, although they are usually not as densely informational as the above excerpt from a course catalog. For example, the following excerpt from an anthropology textbook (Sample 7.4) is somewhat similar in topic and purpose to Text Sample 7.3 above (from the institutional program description): both passages introduce students to anthropology. However, the linguistic features employed for this purpose are strikingly different in the two passages, with the textbook showing a much greater reliance on clause features (many more main verbs and therefore shorter clauses) and a much lesser reliance on complex noun phrase structures.

> **Text Sample 7.4:** Textbook (Anthropology; lower division; tbant1a.fpa) [nouns are in **bold underlined**; verbs are in *italics*]
>
> Learning About the **Past** – The **Material Record**.
>
> This **book** *focuses* on the **human past**, but how do we *learn* about the **past**? How do we *collect* and *analyze* **data** about the ancient **past** of our **species**? In this **book** our **approach** to *understanding* the **human past** will *be* through the **field** of **anthropology**, *defined* properly as the **study** of **humanity**. If you *think* about it, though, nearly all the **courses** you *are* now *taking deal* in some **fashion** with **people** or their **works**. What *makes* **anthropology** different?
>
> The **Anthropology** of the **Past**: **Archaeology** and Physical **Anthropology**.
>
> Many **people** *have* some very strange **ideas** about what **archaeology** and physical (or biological) **anthropology** *are* and what **scientists** in these **fields** *do*. Some **people** *think* **archaeologists** *study* **dinosaurs**. (They *have seen* too many **episodes** of "The **Flintstones**.") In **reality**, the **dinosaurs** *became* extinct more than 60 million **years** before even our earliest **human ancestors** *appeared* on the **scene**. Thanks, at least in **part**, to such **movies** as the **Indiana Jones series**, many *think* that **archaeologists** *are* tough, globetrotting **vagabonds** who *loot* **sites** for **treasure**. Physical **anthropologists** *are* often *stereotyped* as those who *identify* the skeletal **remains** of dead **people**. The most common **question** we *get*, even from **university colleagues**, *is*, "*Dig* up any interesting **bones** lately?" Actually, **archaeology** *is* simply that **branch** of **anthropology** *focusing* on the **human** cultural **past**.

Interestingly, classroom teaching is much more similar in its Dimension 1 characteristics to other spoken registers, including study groups and service encounters, than it is to written academic registers like textbooks. It is rare for an instructor in an American university to lecture for an entire class period, but there are certainly long monologic turns, as the instructor explains a concept or develops a point of view. However, even these extended turns are often highly involved, making dense use of the positive Dimension 1 features with minimal reliance on the highly informational features associated with the negative pole of Dimension 1. For example:

> **Text Sample 7.5:** Classroom teaching (Humanities; Rhetoric; graduate; humrhlegrli004) [selected positive Dimension 1 features are in **bold underlined**]
>
> Instructor: **I think** some of **us feel sort of** really caught in a bind between agency and acculturation. **sort of** um, because **you know I think** a lot of **us** do **want** to use writing, use literacy to um, say **what we want** to say and to help other people say **what they want** to say but at the same time **I think** um, **we're** caught **because we**, **I think we're** questioning **well**, **well you know**, if, if **we**, if **we** teach X-genre **are we promoting** it? if **we don't** at the same time question it and dismantle it and **kind of** take it apart and look at it, and are there, are there other ways? every established genre, every approach, every way that **everything that's** accepted for all of those things that seems like **there's**, there are things that are rejected, not done, not looked at, and as **you were** just **pointing out**, some of the ways that particular systems get promoted are not exactly ethical or correct or at the very least not **everybody** has access to being able to promote certain kinds of discourses and systems, **so I guess**, **I don't know whether** this is really a question or just a comment though **I think we're**, **think** a lot of **us feel** very **sort of** trapped between buying into this is **how** it is and **since it's** this way, **let's** do the best **we** can with it and **sort of saying**, **who's** got the agency **here**? **you know**, **how** is the system beneficial for people? **who** is it hurting?

7.5.2 Dimension 2: Procedural vs. content-focused discourse

In contrast to the spoken-written split identified by Dimension 1, Dimension 2 cuts directly across the spoken/written continuum. Figure 7.2 shows that the registers with large positive scores on this dimension all deal with the rules and procedures expected in university settings, whether in speech (classroom management, service encounters, and office hours) or in writing (course management and institutional writing). At the negative extreme, we find only written academic registers – course packs and textbooks – which have an almost exclusive focus on informational content. Classroom teaching and study groups have intermediate scores on this dimension.

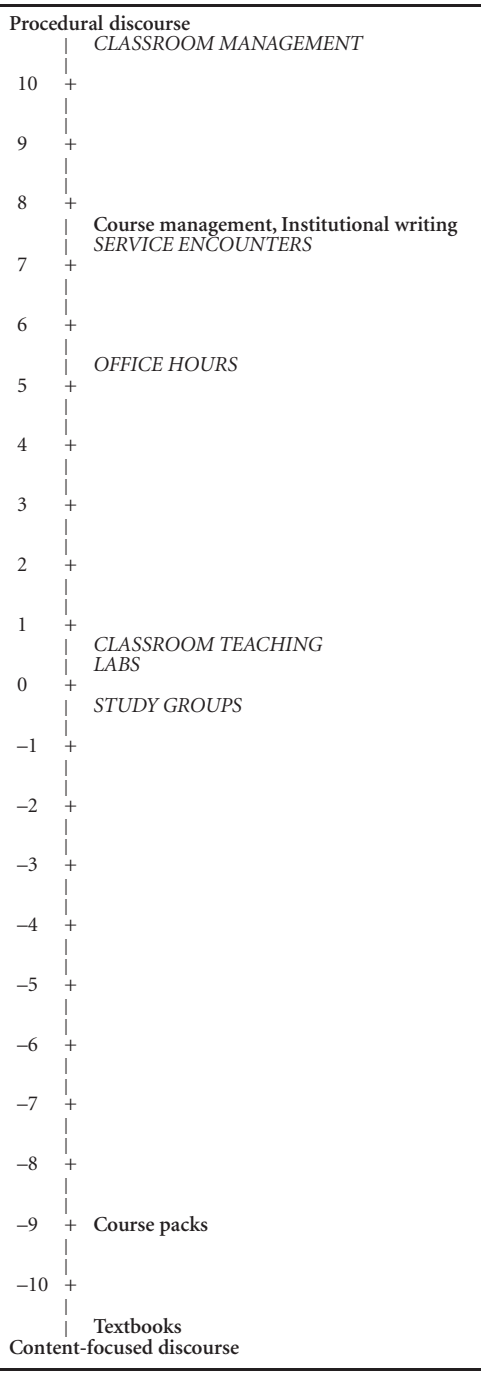

Figure 7.2 Mean scores of registers along Dimension 2 – 'Procedural vs. content-focused discourse'

Considering both the co-occurring linguistic features, together with this distribution of registers, the interpretive label 'procedural vs. content-focused discourse' can be proposed for this dimension. Table 7.1 shows that the linguistic features associated with 'procedural discourse' include necessity and prediction modal verbs, 2nd person pronouns, causative verbs, *to*-clauses with verbs of desire (e.g., *want*), and conditional adverbial clauses. These features are most common in spoken classroom management:

> **Text Sample 7.6:** Classroom management (Humanities; History; upper division; humhicmud_n070) [Positive Dimension 2 features are in **bold underlined**]
>
> OK so, after test number one, **you** can, pretty much push everything out of **your** mind and [3 sylls] and absorb everything for that time and then start over again, OK, although certainly **if you** understand concepts from the first section it **will help you** understand concepts in the second section. These are blue book exams **you'll need to** bring a bluebook, they are mostly essay. uh, one other thing I stated last semester I'm **gonna** do again is that I'm also **gonna** give **you** a take home essay. The take home essay **will** be part of every exam, I**'ll** give **you** the question a week before the exam, **you'll have to** take it home, write it up, type it, and bring it in on the day of the test. it **will** come – it **will** amount to thirty points of **your** one-hundred point exam.
> [...]
> um, let's see, **if** a student misses more than one week of classes **you** should talk to me immediately, **if** you know **you**'re **gonna** be gone. let's say for example **you**'re **gonna** go to Montana for a couple of days this week or something like that **you** might **let** the instructor know **you**'re **gonna** be gone. uh, **if you**'re, I had a woman who was pregnant one semester and she, said well I'm **gonna** be missing part of the class and I said yeah, I think you probably **will** be. OK, but **let** me know. um, **you should let** me know **if you** miss more, **if you** miss a test, **you'd have to** bring me some type of written evidence as to why **you** were gone, just so that it's fair for everybody so that they don't **have to** deal with a whole lot of excuses.

These same features are common in written course management materials. Thus, compare the above excerpt to the following examples from course syllabi:

> At the end of each chapter, **you will** be assigned a series of problems to **help you** write a Chapter Summary. The purpose of the Chapter Summary problems is to **help you** pull together the main ideas of each chapter...
>
> **If you** miss class for two consecutive weeks, **you will** be dropped.
>
> **You will need to** access available resources to find answers to **your** questions and be willing to ask when **you** can't find them. **You will** find that many issues have answers which are complex or ambiguous.

One difference from spoken classroom management is that written course management intermixes personal directives to the individual student (*you*) with impersonal general directives to *students*; for example:

> ***Students*** **will need to** visit the bookstore …
>
> **You** should answer the questions at the end of each chapter to test **your** knowledge of the material, even though there will be few formal assignments. ***Some students*** will have to devote more time to their studies and attend office hours more frequently if ***they*** expect to perform adequately.

In large part, the negative features on Dimension 2 represent technical vocabulary, with 'rare' words from all four content classes (adjectives, nouns, adverbs, and verbs; see Table 7.1). These are all content words that occur in only one text in the T2K-SWAL Corpus. Other negative Dimension 2 features include simple occurrence verbs (e.g., *become, happen, change, decrease, occur*), probability verb + *to*-clause (e.g., *seem / appear to…*), and size adjectives (e.g., *high, large*). The dense use of these co-occurring features is restricted to the written academic registers in the T2K-SWAL Corpus. Text Sample 7.7 illustrates these features in a graduate-level chemistry textbook:

> **Text Sample 7.7:** Textbook (Natural science; chemistry; graduate; TBCHM3.gns)
>
> CHAPTER 11: THE PROTON MAGNETIC RESONANCE SPECTRA OF ORGANIC MOLECULES
>
> Up to now we have been concerned with the magnetic resonance of a single nucleus and with explaining the physical basis of an nmr experiment. We will now turn our attention to the nuclear magnetic resonance spectra of organic molecules and in so doing will encounter two new phenomena: the chemical shift of the resonance frequency and the spin-spin coupling. These two phenomena form the foundation for the application of nuclear magnetic resonance spectroscopy in chemistry and related disciplines.
>
> 1. THE CHEMICAL SHIFT.
>
> The hypothetical spectrum of dimethyltrifluoroacetamide presented at the end of Chapter 1 may have suggested that nmr spectroscopy is employed for the detection of magnetically different nuclei in a compound. For at least two reasons this is not the case. Firstly, experimental considerations make such an application difficult, if not impossible, since conditions and techniques must be modified to measure the resonance frequencies of different nuclei. Secondly, the elemental composition of organic compounds can be determined far more easily and accurately by other techniques such as elemental analysis or mass spectrometry.
>
> …

> The following general statements can be made. For aliphatic C-H bonds the shielding decreases in the series CH3 > CH2 > CH. While the protons of methyl groups at saturated centres absorb at 6 0.9, the resonance for the protons of cyclohexane occurs at 6 1.4. An exception is observed in the case of cyclopropane, the protons of which absorb at 6 0.22. For olefinic protons, the resonances lie in the region from 6 4.0 to 6.5, and only in special instances, such as with compound like acrolein (CH2=CHCHO), below 6 6.5. The resonance signals of protons in aromatic molecules occur in a characteristic region between 67.0 and 8.0. Although Sp2-hybridized bonds are present, as in the olefins, an additional deshielding obviously exists here.

Classroom teaching has an intermediate score on this dimension. In part, this distribution reflects the real time production circumstances of classroom teaching, making it difficult for instructors to use technical/rare vocabulary to the same extent as textbook writers (see the discussion in Chapter 3). In addition, this distribution probably reflects the fact that management talk and content-focused academic talk are less sharply distinguished in classroom teaching than in textbooks.

7.5.3 Dimension 3: Reconstructed account of events

Dimension 3 is in part associated with a narrative orientation, reflected by features like 3rd person pronouns, human nouns, communication verb + *that*-clauses, and past tense (see Table 7.1). However, these features co-occur with a number of other features that express epistemic stance, including likelihood verb + *that*-clause (usually verbs expressing uncertainty, such as *assume, believe, doubt, gather, guess, imagine, seem, suppose, think*); mental verbs (e.g., *think, remember, understand*); mental nouns (e.g., *assumption, attitude, feeling, idea, opinion*); and epistemic stance noun + *that*-clause (e.g., *conclusion, fact, assumption, claim, feeling, idea, impression, opinion, possibility, suggestion, suspicion*). *That*-omission also co-occurs with these features, suggesting that they are usually used in colloquial rather than formal registers.

Figure 7.3 shows that the distribution of registers along Dimension 3 is strongly associated with the spoken versus written modes (similar to Dimension 1): spoken university registers are consistently more 'narrative' than related written registers. Interestingly, the management registers are the least 'narrative' within each mode. Study groups and office hours are especially marked for the use of positive Dimension 3 features.

Consideration of texts from these registers shows that Dimension 3 features are often used to recall what the course instructor said in class. Examples like the

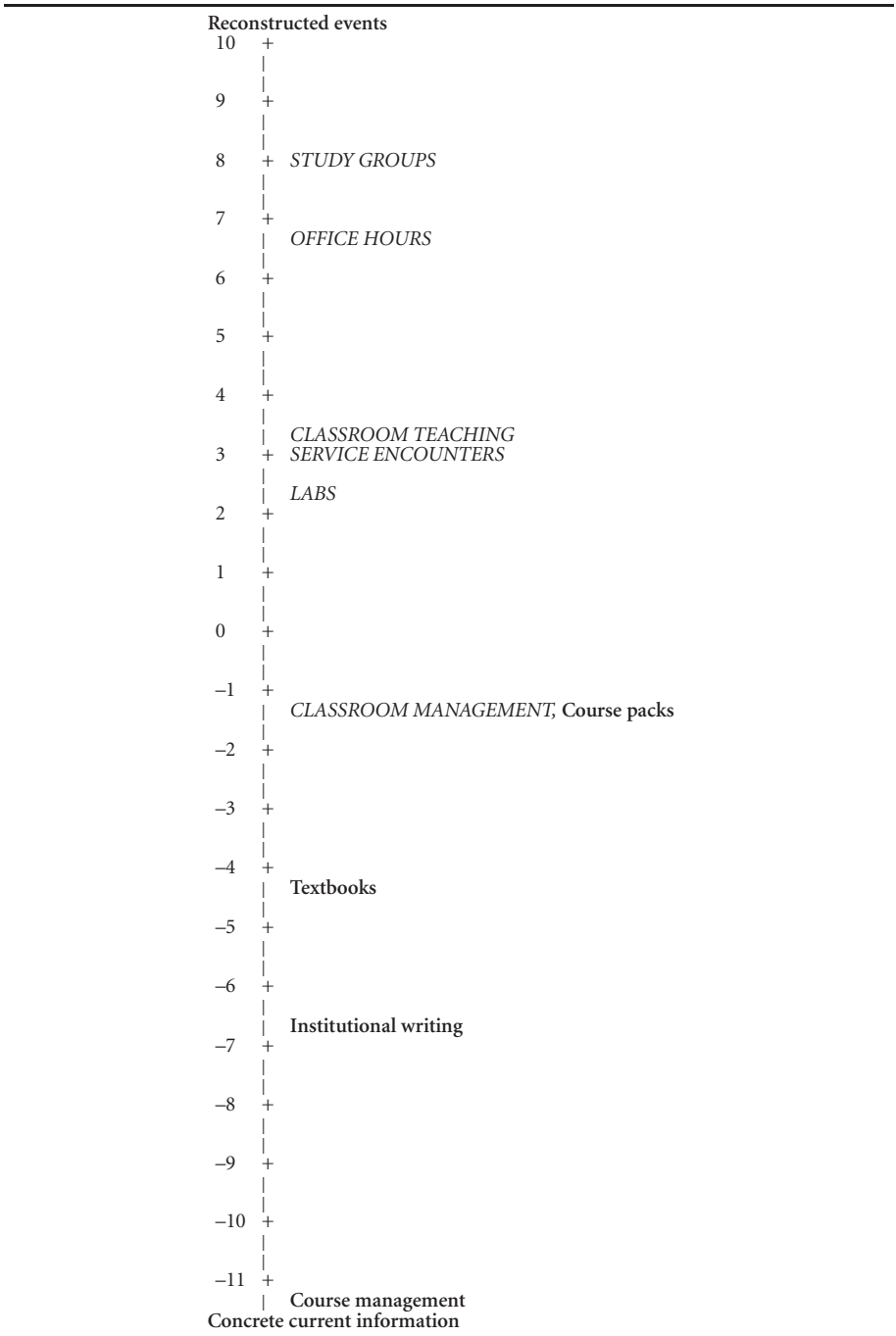

Figure 7.3 Mean scores of registers along Dimension 3 – 'Reconstructed account of events'

following occur frequently in study groups (selected Dimension 3 features are **bold underlined**):

> I **thought he said** it wasn't supposed to change unless you changed it.
>
> Like .. **she said** like the only thing we **went** over in quantum mechanics is Heisenburgs theory and Schroedener's equation. Those er- Schroedener's equation those are the only two things that we **covered** in quantum mechanics. That's all **she said** we need to **know** out of it. She said the rest of it is covered like in organics.

Similarly, in office hours instructors recall what they said in class:

> so I **mean** here's here's what I what I **did** is I just set up the simplex method and I **said** well you **know** the simplex method would normally have you go uh –
>
> Instructor: mhm, uh the key word here was uh uh drilling. so you'd still have to drill to put in a 'pazometer' (sp?).
> Student: right, but I didn't **know** 'pazometer' was considered a well because it, you never **said** there was a pump in it. you just **said** it was like a soda straw. and it was only open at the bottom.
> Instructor: yeah again, the key word there was drilling.

In study groups, it is also common to find students negotiating with one another, trying to reconstruct course content without direct reference to a specific class or what the instructor specifically said. Stereotypical narrative features are used to report past events and situations, while epistemic stance features are used by students to indicate varying degrees of (un)certainty about their knowledge. The following example illustrates a discussion of this type:

> **Text Sample 7.8:** Study group (socposgudpn173) [positive Dimension 3 features are marked in **bold underline**]
>
> 1: Uh in fact as far as, the three major religions, Muslim, Christian and Jew? The Islam are the most –
> 2: Tolerant.
> 1: Tolerant of all the three. So there **was** probably quite a few, in uh, in all the middle eastern countries for that matter. Because –
> 2: All for, the whole two thousand years?
> 1: Yeah because um, **they weren't they weren't** as discriminated against as **they were** in Europe and other countries so **they** –
> 2: OK.
> 1: All **they**, **everybody had** to pay tithes. See that's the other myth about Islam is **they** say, well it's the theocracy and **they** make you pay ten percent of your money.

2: Oh I **remember** – **they, they created** a um, what **did**-what **were they** called those uh, villages those districts with Jews and Christians? In the time of Mohammed? All the way through I can't **remember** what **they**'re called.
1: Mhm. But the point is **they**, **they** um, **did**n't tax. So, in this country if you come and you make money, you pay twenty percent tax.
2: Something like that.
1: In the middle east, in Islam countries you go and you make money and you pay a ten percent tithe, and that's wrong. But it's OK to pay a twenty percent tax. See what I **mean**?
2: Oh yeah I **remember** that.
1: It's like um.
2: I **forgot** about that . . well I **was thinking** also um, that there might have been more Jews uh, during British occupation because they might have immigrated from uh, Great Britain?
1: Mhm I **suppose** it, you **know**, I don't **know**

. . .

2: Well that's the **impression** I got, that that, the Israelis –
1: **They were** the, **they were** the legal nation state, and pretty much, with the backing of the U.S. government, in the, and the British **went** in there and **took** it away from them and, **put** Is-the Jews in there.
1: And **started** Israel.
2: Right. Well my **understanding** is that the Israelis are pretty much the primary aggressors. And they uh, well of course they **took** Jerusalem, I can't remember **was** it sixty seven? Something like that?
1: Sixty seven **was** like, one of the major wars.
2: Yeah that **was** something else though. I'm trying to **remember**.
1: The six day war **was** in like seventy three or something.
2: Yeah.

The negative pole of Dimension 3 includes only three linguistic features: common nouns with concrete/tangible reference (inanimate objects that can be touched, such as *phone, picture, truck, newspaper, handout, syllabus, modem, book, handbook, paper, computer, textbook*); common nouns with concrete but technical reference (tangible objects that are not normally perceived and/or cannot normally be touched, such as *bacteria, electron, sample, schedule, software, solution, margin, virus, mark, internet, message, paragraph, sentence, poem, chapter, equation, exam, statement, diagram*); and common quantity nouns, which relate to quantities, amounts, or durations (e.g., *frequency, future, semester, number, amount, week, month*).

Written course management (syllabi, assignments, and exams) has by far the largest negative score on Dimension 3 (see Figure 7.3). This extreme score reflects a dense use of these negative features coupled with the near total absence of positive

Dimension 3 features. Text Sample 7.9 illustrates these features in an engineering course assignment:

> **Text Sample 7.9:** Written course management (course assignment; engineering; lower division; CMENG1.ASN) [negative Dimension 3 features are marked in **bold underline**]
>
> Overview: There are two **parts** to this **project**. The first **part** is the practice work in **Chapter** A and **Chapter** B where you created the **pages** nomad.htm nomad2.htm and nomad3.htm. The second **part** consists of four steps, where you create a **web page** called project3.htm.
>
> The work that you do as you are reading through the **chapters** in the HTML **text** is practice. Therefore, it will be graded for being attempted, not for being 100% accurate. The **files** you created as you worked through **Chapter** A (nomad.htm) and **Chapter** B (nomad2.htm and nomad3.htm), must be uploaded to your **web directory** on your H drive. Do not put an index.htm **file** in this **directory**.

7.5.4 Dimension 4: Teacher-centered stance

Finally, Dimension 4 seems to be associated with academic styles of stance. The linguistic features defining this dimension include the range of stance adverbials (certainty, likelihood, and attitudinal) and *that*-clauses controlled by stance nouns (e.g., *the fact that*...). *That*-relative clauses and lexical bundles with 'referential' functions (preposition-initial and noun-initial) co-occur with these stance features. Figure 7.4 shows that these features are used primarily in the instructor-controlled spoken registers: classroom management, classroom teaching, and office hours. All written university registers are characterized by the relative absence of these features, as are the student-centered spoken registers (labs, study groups, and service encounters). Following are several examples of these academic stance features in classroom management:

> Instructor: **actually** while I finish the outline, let me pass out the uh **something I'd like you to uh look over** here real quick and sign for me – that you acknowledge **the fact that** you've read and understand the syllabus.
>
> Instructor: January eighteenth **of course** we don't have class. What day is that?
> Student: [unclear]
> Instructor: it's also my birthday, I **always** think that we're taking off on my birthday. uh, but if you link on the jazz home page, you can, there are, there's **actually** jazz music from the twenties
>
> Instructor: all right for the remaining writings, when you take test one – **probably** the second week after spring break on, Tuesday, not Tuesday
> Student: Thursday

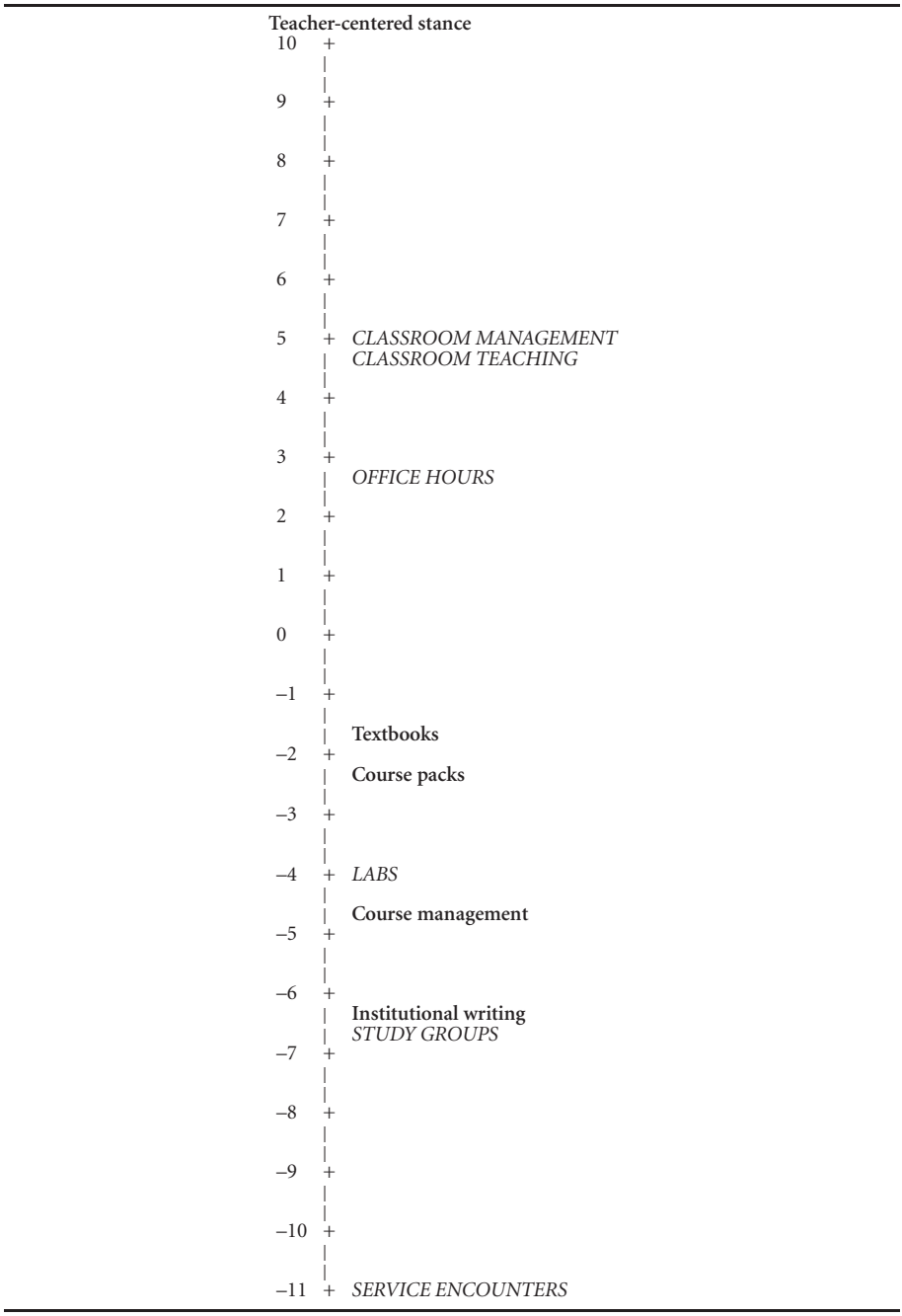

Figure 7.4 Mean scores of Registers along Dimension 4 – 'Teacher-centered stance'

> Instructor: Thursday. you're going to have I think fifteen items, comparable to fifteen of the next twenty three or twenty four [unclear]. so what you have to understand when we go over these items. know which one is correct. and why. make sure you know which one is correct and why. they won't be the exact wording of these but, **certainly** very comparable wording.
>
> Instructor: Let's tabulate those tomorrow, too. Let's do this. Quite **possibly** none of these will be **entirely** satisfactory. [. . .] So **those of you that** have the book, you can, well I guess we've gotta wait. **Those of you that** don't have the book you can do it now.
>
> Instructor: quickly now – the department came down and I know, Mark's been working on those so, **hopefully**, first of next week, at the latest we ought to have it up and working. **kind of**, continuing the tradition that has gone on in the past, several semesters or at least the past semester. I, am following Professor Anderson's lead on doing a number of things – **one of the things that** he did, was have, uh, Professor Brown from the English, department come over and make a presentation. uh, last year, and [er] last semester. I invited her back she **gracefully** agreed to do that and she's going to, talk to us about, can we be a good engineer if you can't communicate?

The same combination of academic stance features associated with Dimension 4 is common in classroom teaching, as in:

> Instructor: uh what they have done here if you don't have this, I **really** can't write the whole thing out but uh **hopefully** by, you know, just going through your notes you'll be able to relate this to the problem. they give us a linear programming problem and then they say which of the following is the initial tableau for the whole problem. now uh I think there **probably** could be more than one way of getting there, but only one of these will be correct.
>
> . . .
>
> we're going to have a less than or equal to constraint for the second constraint, but the third constraint will be an equal to constraint. so **one of the things that** we can do is uh I mean **actually** this **maybe** gets us almost where we need to be.
>
> Instructor: um well **actually** it's two questions. one how like I know **a lot of people that** are **really really** good at their job? and they're **really** not interested in telling people about it and about how to do it.

7.5.5 The distribution of stance features across the university dimensions

It is interesting to note that three of the four dimensions in this analysis are strongly associated with spoken/written differences, and stance features are important defining characteristics for all three dimensions. Dimension 1 is the most general spoken/written distinction, with all spoken registers having posi-

tive scores, and all written registers having negative scores. This dimension shows that all spoken university registers are characterized by a focus on personal stance, represented by:

– *that*-clauses controlled by likelihood verbs (e.g., *think*),
– *that*-clauses controlled by certainty verbs (e.g., *know*),
– *that*-clauses controlled by communication verbs (e.g., *said*),
– certainty stance adverbials (e.g., *actually*),
– hedges (e.g., *kind of, something like*),
– likelihood stance adverbials (*maybe*),
– lexical bundles consisting of clause segments, which often serve stance functions:
  pronoun-initial (*I don't know if*)
  WH-word initial (*what do you think*)
  verb initial (*want to talk about*).

These stance features on Dimension 1 occur together with features of interaction and high involvement (e.g., contractions, 1st and 2nd person pronouns, questions, etc. . .), reflecting the general discourse styles and priorities of spoken university registers.

Dimension 3 also includes several stance features:

– *that*-clauses controlled by communication verbs (e.g., *said*),
– *that*-clauses controlled by likelihood verbs (e.g., *think*),
– *that*-clauses controlled by stance nouns (e.g., *fact*).

These features co-occur with features reflecting a general narrative orientation, functioning as speakers' attempts to reconstruct past events. Similar to Dimension 1, Dimension 3 features are common in most spoken university registers; they are especially prevalent in 1-on-1 interactions about academic topics: study groups and office hours.

Several of the same stance features that occur on Dimensions 1 or 3 are also found co-occurring on Dimension 4:

– certainty stance adverbials (e.g., *actually*),
– attitudinal stance adverbials (e.g., *amazingly*),
– likelihood stance adverbials (e.g., *maybe*),
– *that*-clauses controlled by stance nouns (e.g., *fact*).

However, in this case these features co-occur with 'informational' features, such as *that* relative clauses and preposition-initial lexical bundles. Dimension 4 features are common in several spoken registers, while all written university registers are marked by the relative absence of these features. However, Dimension 4 also distinguishes among different kinds of spoken university registers: only the

teacher-centered spoken registers – office hours, classroom teaching, and classroom management – are marked for the dense use of the academic stance features associated with Dimension 4. In contrast, student-centered registers, even when they focus on academic topics (as in study groups), are marked by the absence of these co-occurring features.

Dimension 2 is also defined in part by stance features, especially modal verbs (necessity and prediction) and verbs of desire+*to*-clause constructions expressing procedural/directive functions. Unlike the other three dimensions, the stance features grouped on Dimension 2 are associated with both speech and writing, but restricted to the 'management' registers.

Thus, although they are all associated with the expression of stance, these four dimensions reflect very different functional considerations, and as a result, they distinguish among different sets of registers:

Dimension	**1**	**2**	**3**	**4**
Functional interpretation of positive features	stereotypically 'oral'; interaction and high personal involvement	procedural / directive discourse	collaborative reconstruction of past events	academic stance
Registers marked for these stance features	ALL spoken university registers	spoken and written management registers	1-on-1 spoken academic registers, especially study groups and office hours	teacher-centered registers: classroom management, classroom teaching, and office hours

7.5.6 Differences among academic disciplines

Two of the university dimensions are also associated with important differences across academic disciplines: Dimension 2 ('Procedural vs. content-focused discourse'), and Dimension 3 ('Reconstructed account of events'). Figures 7.5 and 7.6 plot the dimension scores for each discipline in classroom teaching and textbooks. The distributions are highly systematic along both dimensions, with disciplines following the same pattern within both textbooks and classroom teaching.

Figure 7.5 shows that Dimension 2 ('Procedural vs. content-focused discourse') distinguishes sharply between the two technical disciplines: Engineering is 'procedural' in orientation (in both speech and writing), while natural science is by far the most 'content-focused'. Business and education also show a strong tendency to favor procedural styles. These differences are found in both classroom teaching and in textbooks.

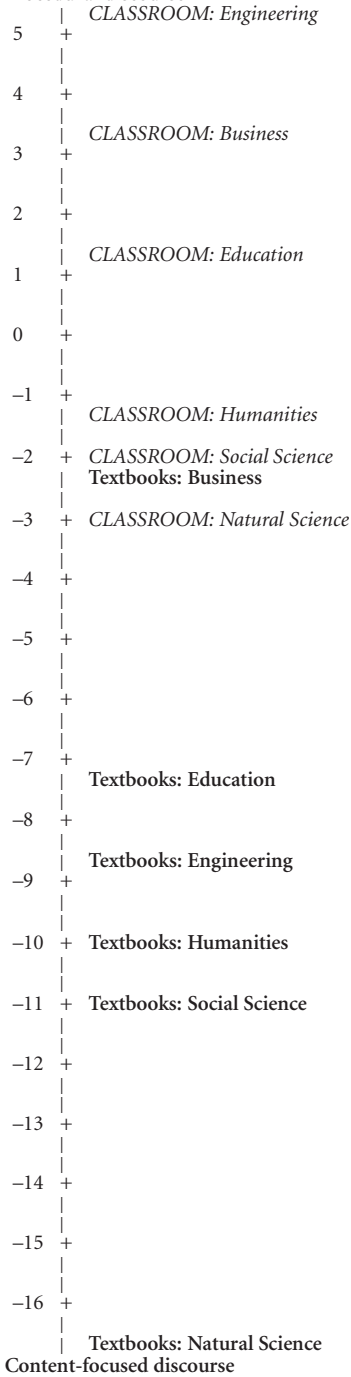

Figure 7.5 Mean scores of disciplines along Dimension 2 – 'Procedural vs. content-focused discourse'

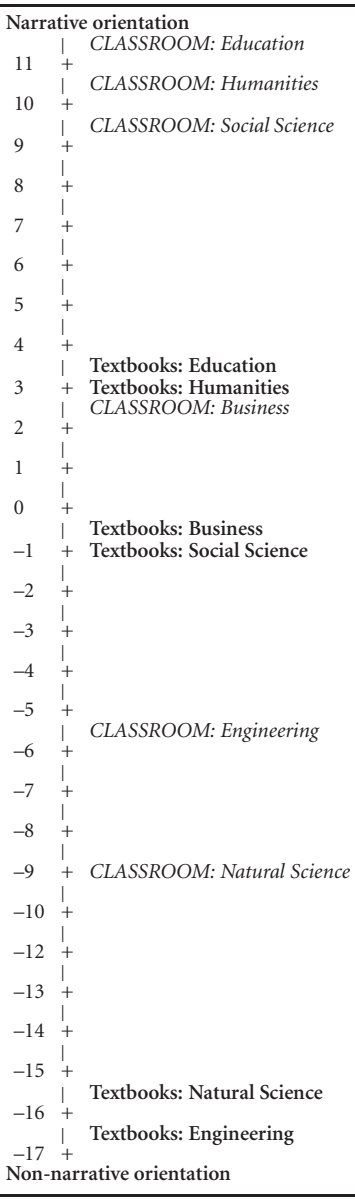

Figure 7.6 Mean scores of disciplines along Dimension 3 – 'Reconstructed account of events'

In engineering classroom teaching, procedural discourse is part of normal class lessons, where students are asked to think through the logical steps required to solve real-world problems. For example:

Text Sample 7.10: Classroom teaching (engineering; upper division; engceleudli029) [positive Dimension 2 features are marked in **bold underline**]

And **you**, one of the things that seems to me **you'd wanna** look at very quickly is, steam is very expensive, raw material basically **you**'re **using** steam [5 sylls]. When **you**, convince the steam, to appear somewhere, uh, first of all **you**'ve **built** the energy into generating that steam and then **you**'re venting it out here and, it's not, it vents it out just not pure enough to go to the boiler feed water. **You**'ve **got**, [3 sylls], so **you** now **you have to** treat it **if you're going to** recycle it.

Business and education classroom teaching also commonly use the "procedural" features associated with Dimension 2:

Text Sample 7.11: Classroom teaching (business; upper division; busacleudms058) [positive Dimension 2 features are marked in **bold underline**]

We **have to** determine basis for gains, basis for loss, and we're **gonna put** a rate for depreciation in there as well. Now the reason I **want to let you help** me write these rules in, when **you** have a dual basis rule that is a basis for gain and a basis for loss there's only so many ways **you** can describe it. OK once **you** know what this is, this **has to** be similar to it.

Text Sample 7.12: Classroom teaching (education; upper division; edubeleudhn133) [positive Dimension 2 features are marked in **bold underline**]

I think that Alb offered some suggestions in the book that first **you need to** preview the book before you read it to the class. uh, Second, perhaps **you need to** sit down with the other teachers that you plan with on a regular basis, and see how the book really fits into your current thematic unit. [. . .] I'm, um, I'm **going to give you** some information now about selecting multi-cultural textbooks. And, **you** know I think we **need to** be sensitive to the population of students in the classroom. And, I'm **going to give you** some examples of choosing books for Native Americans.

Textbooks are generally much less marked for 'procedural' features than classroom teaching. However, we find similar differences among the disciplines. Business textbooks are the most marked for the use of procedural features; for example:

Text Sample 7.13: Textbook (business; upper division; tbmkt2.cmr) [positive Dimension 2 features are marked in **bold underline**]

After specifying the basic information that **will** be sought, the researcher **needs to** specify how it will be gathered. [. . .] **If** the researcher decides on a disguised unstructured questionnaire in which subjects **will** be **shown** a picture and asked to tell a story about it, a telephone interview **would** be out of the question [. . .]

> **If** researchers decided to **use** a lengthy attitude scale, for example, they **would** probably **have to** rule out telephone interviews. [. . .] Thus, the researcher **must** specify precisely what primary data are needed, how these data might be collected, what degree of structure and disguise **will** be **used**, and then how the questionnaire **will** be administered.

Education and engineering textbooks also use these features to a greater extent than the other disciplines:

Examples from education textbooks:

> One of the best teaching manuals says **you should** ask a question first, then name the child **you want to** have answer it.
>
> Do not make any rule that **you** are not willing to enforce every time it is broken. [. . .] **If** the behavior in question is not serious enough for **you** to enforce every time, the rule is not worth having. **If you fail to** enforce one of your classroom rules, **you** have just taught the entire class that your rules are hot air.

Examples from engineering textbooks:

> **If you apply** the phase rule to a multi-component gas-liquid system at equilibrium, **you will** discover that the compositions of the two phases at a given temperature and pressure are not independent.
>
> To design a separation process unit, **you must** know certain physical properties of the system with which **you will be working**. Before **you** can design a distillation column, for example, **you must** know how volatile each feed component is [. . .]

At the opposite extreme, natural science textbooks are characterized by a heavy reliance on technical vocabulary and 'rare' words, coupled with the absence of 'procedural' discourse features. Text Sample 7.7 (in Section 7.5.2 above) illustrates these characteristics in a chemistry textbook.

In contrast to their opposite characteristics along Dimension 2, Figure 7.6 shows that engineering and natural science are similar to each other in favoring non-narrative styles along Dimension 3, in both speech and writing. At the other extreme, education, humanities, and social science are all much more likely to incorporate reconstructed accounts of past events. These features are most prominent in classroom teaching, but textbooks from the same disciplines are also marked for the relatively frequent use of narrative features. Text Sample 7.14 illustrates these characteristics from history classroom teaching, as the instructor tries to reconstruct historical settings and events in response to a student question:

> **Text Sample 7.14:** Classroom teaching; History (upper division; humhileudmn089) [past tense verbs and 3rd person pronouns are **bold underlined**; other positive Dimension 3 features are in ***bold italics***]

Student: But why would, I ***mean***, why would, uh, China want to throw out Soviet technocrats [. . .]
Instructor: And I ***think*** that's, you ***know***, that's like the key issue there in, is absolute [2 sylls] relationship as it develops that, that it's not one [2 sylls] process. [. . .] OK? Um, and-and-uh, so what we have here, of course one can face Mao's dissatisfaction to some extent, the very very early roots of Mao's dissatisfaction with, uh, ideological dissatisfaction, I'm not ***talking*** just about personal dis-dissatisfaction, but from **his** ideological point of view from the way **he interpreted** history is toward the development, social development and the revolution particularly. From that view point, **his** earliest dissatisfactions with the, uh, leadership of the Soviet Union, with Stalin's leadership of the Soviet Union, will go all the way back to the nineteen thirties. I ***mean*** **he**, **believed**, even then, that there **was** not a clear understanding of China's situation, China's revolutionary situation, on the part of Stalin and the Soviet nation. Whereas Mao **was**, even back in the nineteen thirties **he was** already formulating that [2 sylls] strategy. Formulating that strategy which **was called**, you ***know***, countryside surrounding the cities. That's a way to conduct revolution . At that same time, **he was** already mindful of the ***fact that*** the Soviet leadership **was** of a different kind. Standing from the ***fact that*** the Soviet experience itself **was** by and large an urban revolution, experience. The workers in Moscow, you ***know***, in the Soviet Union the peasants **had** consistently had very little, a small role in the revolution. More of an exploited role than an active role . And Mao **had** already kind of seen that distinction in the ideological between **himself** and the Soviet leadership in terms of ideology of revolution.

In education classroom teaching, positive Dimension 2 features often convey personal narratives coupled with personal commentary, rather than being used to reconstruct past historical events. For example:

Text Sample 7.15: Classroom teaching; Education (graduate; edubelegrhn161) [past tense verbs and 3rd person pronouns are **bold underlined**; other positive Dimension 3 features are in ***bold italics***]

Instructor: [. . .] I'm sorry I **interrupted** someone here. Um. Jenny?
Student: [unclear]
Instructor: No go ahead. Go ahead Jenny.
Student: Um, [8 sylls] I **came** here when I was thirteen and I **went** to public school in California, and I **went** [3 sylls] to school, my, teacher **did**n't, as you ***know*** [4 sylls] care when I, I, **did**n't [3 sylls] pronunciation [6 sylls] thirty to thirty five kids in a class. And then, I **went** to college and then, after college I **was** ***thinking*** oh my God my mom, [3 sylls] for ten years, and, I **did**n't get anywhere. [6 sylls], Mom [1 syll], and always [8 sylls] on southeast, Asian countries that, [4 sylls], they **came** here and [10 sylls] my sister **went** to college, she has to pay like thousand dollar per, um, quarter, for her tuition. And I **was** ***thinking***, how am I even going,

I don't have any [5 sylls] um, um, education. I ***feel*** I cannot, I should ***say*** this, [7 sylls], for selfish. I ***mean***. Where's my mom [3 sylls] where's my chance [3 sylls]. And I can ***understand*** [8 sylls] opinion and I can ***understand*** those people I ***think*** because, **they** paying the tax, but **they did**n't get anything from, their money. I don't ***know***.
Instructor: OK so we ***see*** a real decline in public support for education that's reflected in, in the tax rates in California. And also just reflected in general opinions of, of California's schools. I **went** through school in California, um, a long time ago. Um, and when I **went** through the University of California system it only cost ninety nine dollars a quarter. Um, in my high school we **had** sixteen national merit finalists and it **was** a public high school. Everyone **did** well, uh, **went** on to college and **did**n't have to pay much for college. But in, in the generation that **ensued** we've ***seen***, a real change in expectations, in attitudes, and just in support for public education. And this is a real crisis that goes beyond um, a simple um, discussion of language in a classroom, or how we're going to, um, work with new immigrants to the California community. And, and I ***think*** these are serious issues that we ***need*** to ***talk*** about, and, and solve.

Textbooks from disciplines with a focus on the past (e.g., history) rely more heavily on narrative discourse, including long narrative sections written entirely in the past. Many lower division history textbooks are written primarily in the past tense, since they simply narrate past events and circumstances. These books sometimes also document the beliefs, expectations, and other attitudes of historical characters (using verb+*that*-clause constructions), but the majority of the text is given to simple narration; for example:

Text Sample 7.16: Textbook; Humanities (history, lower division; tbhis1.bap) [past tense verbs and 3rd person pronouns are **bold underlined**; other positive Dimension 3 features are in ***bold italics***]

Much of the early history of the United States **was** written by New Englanders, who **were** not disposed to emphasize the larger exodus of Puritans to the southerly islands. When the mainland colonists **declared** independence in 1776, they ***hoped that*** these island outposts would join them, but the existence of the British navy **had** a chilling effect.

These common convictions deeply **shaped** the infant colony's life. Soon after arrival the franchise **was** extended to all "freemen" [. . .]

As mentioned above, engineering and natural science texts – from both classroom teaching and textbooks – have large negative scores on Dimension 3. These scores result from the absence of positive Dimension 3 features, coupled with frequent use of the negative features: especially quantity nouns (e.g., *length, amount*) and concrete nouns (including nouns referring to a specific entity but having technical meaning, like *electron*). These negative Dimension 3 scores reflect an interesting

mixture of topics and referring expressions in these disciplines, with highly technical discourse that discusses complex mathematical relationships among everyday concrete entities.

Text Sample 7.17: Classroom teaching; Engineering (Electrical engineering; lower division; engeeleldli015.txt) [concrete nouns and quantity/mathematical nouns are **bold underlined**]

let's continue to look at this problem and then we want to move on. the author says that given these .. the structure he's going to do a field **map**. now this **book** for, unfortunately this **book** doesn't have the method to do **field** mapping so unless [unclear words] you're responsible for doing **field** mapping, uh although it's a fun project, but if you are given a field **map**, you should be able to figure out the **capacity**. you're told, the author says that he's going to do each one of these as **squares** curvilinear **squares**, and so starting out in the center, these are little **squares**, and on each of the little **squares**, there is a equal potential **line** .. and there are flux **tubes**. now what do we mean by flux **tubes**? … we mean that if we take the **integral** of V dot V.S. over the **area** of the flux **tube** we get the same answer everywhere. in that **tube**. now the **integral** of V dot V.S. here if these are nice and square, the **integral** of V. dot V.S. on this top surface, is a flux, remember the (E) **field** is going through here like this … mm, the flux coming down here like that, so is a same flux running through every one of the **tubes**.

Text Sample 7.18: Textbooks; Engineering (Mechanical engineering; graduate; TBMCE3.GVD) [concrete nouns and quantity/mathematical nouns are **bold underlined**]

Although many ride problems are peculiar to a specific **road**, or **road** type, the notion of "**average**" **road** properties can often be helpful in understanding the response of a **vehicle** to **road** roughness. The general similarity in the **spectral content** of the **roads** seen in Figure 5.2 (that **elevation amplitude** diminishes systematically with increasing **wavenumber**) has long been recognized as true of most **roads**.

Consequently, **road inputs** to a **vehicle** are often modeled with an **amplitude** that diminishes with **frequency** to the second or fourth **power** approximating the two linear segments of the **curve** shown in the figure. The **average properties** shown in the figure are derived from recent studies of a large **number** of **roads**.

7.6 Comparison of the general spoken/written dimensions (1988) and the T2K-SWAL dimensions of variation

There are several interesting points of similarity and contrast between the 1988 MD model of register variation and the MD analysis of university registers presented here. Dimension 1 is especially similar in the two analyses: in both cases, Dimension 1 represents a fundamentally important parameter of variation, composed of an extremely large set of linguistic features, and defining a basic opposition between oral and literate registers. However, two interesting points of difference emerge even for this dimension: First, Dimension 1 in the T2K-SWAL analysis actually collapses three major oral/literate dimensions from the 1988 analysis (Dimensions 1, 3, and 5). Second, the mode differences were not as sharply distinguished in the 1988 analysis. That is, while Dimensions 1, 3, and 5 in the 1988 analysis all distinguished between stereotypical speech and writing (e.g., conversation and academic prose), they also showed overlap among other spoken and written registers (such as fiction and personal letters). In contrast, Dimension 1 in the T2K-SWAL analysis shows an absolute difference between spoken and written registers in the university context.

In some respects, Dimension 3 in the T2K-SWAL analysis ('Reconstructed account of events') resembles Dimension 2 in the 1988 analysis ('Narrative discourse'). However, Dimension 2 in the 1988 analysis had an even stronger association with stereotypically narrative styles (e.g., features like 3rd person pronouns, past tense verbs, perfect aspect, and communication verbs), which are especially common in written fiction. As noted in 7.2.2 above, the Biber et al. (2002) study showed that the 1988 Dimension 2 does not discriminate among university registers. That is, no university register is marked for the dense use of the stereotypical narrative features that define the 1988 Dimension 2. In contrast, the T2K-SWAL Dimension 3 incorporates both narrative features and stance features, functioning together in collaborative reconstructions of past events (rather than stereotypical fictional narrative). This dimension is associated with important differences among university registers (including a general distinction between spoken and written university registers), plus important differences among academic disciplines.

The other two dimensions in the T2K-SWAL analysis do not have close counterparts in the 1988 MD analysis. Dimension 4 ('Academic stance') is in some ways similar to Dimension 4 in the 1988 analysis ('Overt persuasion'), but the T2K-SWAL dimension plays a much more important role in distinguishing among university registers. Dimension 2 in the T2K-SWAL analysis ('Procedural vs. content-focused') has no counterpart in the 1988 analysis, indicating a parameter of variation that is distinctively important in the university setting.

Another important difference between the two analyses is the ubiquitous distribution of stance features across all four T2K-SWAL dimensions. In the 1988 MD analysis, stance features are restricted primarily to Dimension 1 ('Involved vs informational production') and Dimension 4 ('Overt argumentation or persuasion'). In contrast, stance features are prominent on all four T2K-SWAL dimensions. In part, this difference is due to the inclusion of additional stance features in the factor analysis for the T2K-SWAL study. However, the difference in the analyses also reflects the variety in the kinds of stance expressed in university registers, as well as a greater reliance on stance features overall. I return to the general importance of stance in university registers in Chapter 8.

Finally, the T2K-SWAL dimensions correspond to important differences among academic disciplines, while there are no important differences between disciplines identified with respect to the 1988 dimensions (see Biber et al. 2002). In particular, important differences were found here with respect to both T2K-SWAL Dimension 2 (engineering, business, education versus natural science) and T2K-SWAL Dimension 3 (education, humanities, social science versus engineering and natural science). Methodologically, these patterns illustrate the utility of carrying out a new factor analysis for a specific discourse domain. That is, the T2K-SWAL factor analysis identified important patterns of linguistic variation among academic disciplines, realized by specific configurations of linguistic features that were not captured in the general 1988 factor analysis. By basing the T2K-SWAL factor analysis on a corpus that represented both register differences and disciplinary differences, the study was able to capture dimensions that reflected the patterns of co-occurrence and alternation across the full range of university varieties.

Notes

1. See Biber (2003b) for a preliminary report on this analysis.

2. The only previous MD analysis of school registers based on a new factor analysis has been Reppen (2001). That study analyzed a corpus of elementary students' spoken and written texts (including both textbooks and essays written by the students). The resulting factor analysis had five primary dimensions. Some of these are similar to dimensions in the general adult (1988) model; for example, Dimension 1 in the student model, which was interpreted as 'Edited informational vs. on-line informational discourse', is similar to Dimension 1 in the 1988 model. However, other dimensions are unique to the student model; for example, Dimension 4 was interpreted as 'Projected scenario', and Dimension 5 was interpreted as 'Other-directed idea justification' (see Reppen 2001: 191–194).