

their language, speakers also have extensive knowledge about the use of their language. The former knowledge is grammatical competence, which includes the traditional areas of phonology, syntax, and semantics. The latter knowledge is known as 'communicative competence', and includes formal knowledge of the range of speech-act variation, dialect variation, and register variation, as well as knowledge of when these different linguistic forms are appropriate. Grammatical competence is concerned with the linguistic structure of 'grammatical' utterances; communicative competence is concerned with the form and use of all language – both speech and writing. Within this framework, neither speech nor writing needs to be considered primary to the exclusion of the other. Rather, both require analysis, and the linguistic comparison of the two modes becomes an important question.

Of course, in terms of human development, speech has primary status. Culturally, humans spoke long before they wrote, and individually, children learn to speak before they read or write. All children learn to speak (barring physical disabilities); many children do not learn to read and write. All cultures make use of spoken communication; many languages do not have a written form. From a historical and developmental perspective, speech is clearly primary.

Once a culture has developed written communication, however, there is no reason to regard writing as secondary within that context. It has long been known that cultures exploit variation in linguistic form for functional purposes. For example, variation between lexical items such as *lorry* and *truck* functions to mark geographical differences; variation between pronunciations such as [ka:] versus [kar] and [ðis] versus [dis] functions to mark social differences; variation in address terms, such as *Dr. Jones* versus *Sue*, functions to mark the formality of the situation and the social role relationship between speaker and listener. Similarly, once a culture develops a written form in addition to a spoken form, the two modes come to be exploited for different communicative purposes. Although either speech or writing *can* be used for almost any communicative need, we do not in fact use the two forms interchangeably. Rather, depending on the situational demands of the communicative task, we readily choose one mode over the other. Usually this choice is unconscious, since only one of the modes is suitable or practical. For example, we have no trouble choosing between leaving a note for someone or speaking to the person face-to-face; the situation dictates the

mode of communication. Similarly, we have no problem deciding between writing an academic exposition for an audience and addressing the audience by means of a spoken lecture. We could in fact write a lecture or a note to a physically present audience, but this would take more effort and time than required, and it would fail to take advantage of the opportunities for interaction. Conversely, speaking a lecture or note to an addressee who is separated by time or place is usually not possible at all; apart from the use of telephones and tape recorders, the written mode is required in situations of this type. These simple examples illustrate the fact that the two modes of communication have quite different strengths and weaknesses, and they therefore tend to be used in complementary situations. From this perspective, neither can be said to be primary; they are simply different. The linguistic characteristics of each mode deserve careful attention, and the relationship between the two modes must be investigated empirically rather than assumed on an a priori basis.

## 1.2 Dimensions and relations

In the present book, spoken and written texts are compared along 'dimensions' of linguistic variation. Researchers have considered texts to be related along particular situational or functional parameters, such as formal/informal, interactive/non-interactive, literary/colloquial, restricted/elaborated. These parameters can be considered as dimensions because they define continuums of variation rather than discrete poles. For example, although it is possible to describe a text as simply formal or informal, it is more accurate to describe it as more or less formal; formal/informal can be considered a continuous dimension of variation.

I will illustrate the concept of 'dimension' in this section by analysis of a few linguistic features in four texts. This illustration greatly oversimplifies the linguistic character of the dimensions actually found in English. Chapters 5–8 present a full analysis based on the distribution of 67 linguistic features in 481 texts. The discussion here thus provides a conceptual description of dimensions, rather than actually describing the complex patterns of variation in English speech and writing.

Following are two quite distinct text samples, which differ along several dimensions. Readers should identify some of the differences between them before proceeding to the following discussion.

## Text 1.1: Conversation – comparing home-made beer to other brands

- A:** *I had a bottle of ordinary Courage's light ale, which I always used to like, and still don't dislike, at Simon Hale's the other day – simply because I'm, mm, going through a lean period at the moment waiting for this next five gallons to be ready, you know.* 1–6
- B:** *mm* 7
- A:** *It's just in the bottle stage. You saw it the other night.* 8
- B:** *yeah* 9
- A:** *and, mm I mean, when you get used to that beer, which at its best is simply, you know, superb, it really is.* 10–11
- B:** *mm* 12
- A:** *you know, I've really got it now, really, you know, got it to a T.* 13–14
- B:** *yeah* 15
- A:** *and mm, oh, there's no, there's no comparison. It tasted so watery, you know, lifeless.* 16–17
- B:** *mm* 18

## Text 1.2: Scientific exposition

- Evidence has been presented for a supposed randomness in the movement of plankton animals. If valid, this implies that migrations involve kineses rather than taxes (Chapter 10). However, the data cited in support of this idea comprise without exception observations made in the laboratory.* 1–5

Text 1.1 is taken from an ordinary, face-to-face conversation between friends. It represents the type of communication that we all experience every day. Text 1.2 is much more specialized, coming from a scientific exposition. In contrast to the conversation, relatively few speakers of English commonly read texts like 1.2, and an extremely small proportion are expected to write texts of this type. We might thus distinguish texts 1.1 and 1.2 on a dimension of common versus specialized.

These texts might also be contrasted on a dimension of unplanned versus planned. In text 1.1, speaker A talks without careful planning. At one point he switches topic in the middle of a sentence – in line 10, he begins a thought with *when you get used to that beer*, and two utterances later, in line 16, he completes the sentence with *there's no comparison*; in between these two utterances he notes that his homemade beer is superb when made properly (lines 10–11), and that he really knows how to make the brew now (lines 13–14). Text 1.2 is quite different, having a very careful logical progression indicating careful planning. An idea is presented in lines 1–2, implications of the idea are given in line 3, and the idea is qualified in lines 4–5. This logical progression continues in the rest of text 1.2.

There are several other dimensions that these two texts could be compared along. For example, text 1.1 is interactive while text 1.2 is not; in text 1.1, speaker A refers directly to himself and to speaker B (*I* and *you*), and speaker B responds to A. Text 1.1 is dependent on the immediate situation to a greater extent than text 1.2; in text 1.1, speaker A assumes that B can identify *Simon Hale's* (line 3), *the other day* (line 3), *this next five gallons* (line 5), and *the other night* (line 8). The speaker in text 1.1 displays his feelings enthusiastically and emphatically, while the feelings of the writer in text 1.2 are less apparent; speaker A in text 1.1 repeatedly emphasizes his point with *really*, *simply*, and *you know* (lines 4, 6, 11, 13, 17).

When only two texts are compared, these parameters seem to be dichotomies. If we add a third text, however, we begin to see that these parameters define continuous dimensions. Thus, consider text 1.3 below:

## Text 1.3: Panel discussion – discussing corporal punishment as a deterrent to crime

- W:** *But Mr. Nabarro, we know that you believe this.*
- L:** *quite*
- W:** *The strange fact is, that you still haven't given us a reason for it. The only reason you've given for us is, if I may spell it out to you once more, is the following: the only crime for which this punishment was a punishment, after its abolition, decreased for eleven years.*

***You base on this the inference that if it had been applied to crimes it never had been applied to, they wouldn't have increased.***

***Now this seems to me totally tortuous.***

Text 1.3 is intermediate between texts 1.1 and 1.2 with respect to the dimensions outlined above. Text 1.3 is certainly not a common everyday communication like text 1.1, but it is not as specialized as text 1.2; text 1.3 is relatively unplanned, but it is more carefully organized than text 1.1; text 1.3 is interactive, but not to the extent of text 1.1; text 1.3 shows little dependence on the immediate situation, but more so than text 1.2; and for the most part, the main speaker in text 1.3 does not reveal his own feelings, although they are more apparent than those of the writer in text 1.2. Text 1.3 is more like text 1.1 with respect to some of these dimensions, and more like text 1.2 with respect to others. However, it has an intermediate characterization with respect to texts 1.1 and 1.2 on each dimension, indicating that these are continuous parameters rather than simple dichotomies.

To this point, we have discussed the notion of dimension from a situational or functional point of view. It is also possible to discuss this notion from a strictly linguistic perspective. In the same way that texts can be described and compared in terms of their situational characterization, there are dimensions that compare texts in terms of their linguistic characterization, e.g., nominal versus verbal, or structurally complex versus structurally simple. Thus consider texts 1.1, 1.2, and 1.3 again. A general impression of text 1.1 is that it is verbal rather than nominal (i.e., many verbs, few nouns) and that it is structurally simple (e.g., little phrasal or clausal elaboration). Text 1.2, on the other hand, seems to be extremely nominal and structurally complex, while text 1.3 seems to have a linguistic characterization between these two. Several questions arise, though: (1) What evidence can we give to support these linguistic impressions? (2) Do these characterizations represent a single linguistic dimension, or two dimensions, or more than two? How can a researcher determine how many linguistic dimensions are required to account for the variation among a set of texts? (3) Are there other linguistic dimensions that are not represented by the above linguistic impressions? If so, how can they be discovered?

I develop an overall empirical approach in the present book that addresses these questions (cf. Section 1.3, Section 3.5, and Chapter 4).

The raw data of this approach are frequency counts of particular linguistic features. Frequency counts give an exact, quantitative characterization of a text, so that different texts can be compared in very precise terms. By themselves, however, frequency counts cannot identify linguistic dimensions. Rather, a linguistic dimension is determined on the basis of a consistent co-occurrence pattern among features. That is, when a group of features consistently co-occur in texts, those features define a linguistic dimension. It should be noted that the direction of analysis here is opposite from that typically used in studies of language use. Most analyses begin with a situational or functional distinction and identify linguistic features associated with that distinction as a second step. For example, researchers have given priority to functional dimensions such as formal/informal, restricted/elaborated, or involved/detached, and subsequently they have identified the linguistic features associated with each dimension. In this approach, the groupings of features are identified in terms of shared function, but they do not necessarily represent linguistic dimensions in the above sense; that is these groupings of features do not necessarily represent those features that co-occur frequently in texts. The opposite approach is used here: quantitative techniques are used to identify the groups of features that actually co-occur in texts, and afterwards these groupings are interpreted in functional terms. The linguistic dimension rather than functional dimension is given priority.

This approach is based on the assumption that strong co-occurrence patterns of linguistic features mark underlying functional dimensions. Features do not randomly co-occur in texts. If certain features consistently co-occur, then it is reasonable to look for an underlying functional influence that encourages their use. In this way, the functions are not posited on an a priori basis; rather they are required to account for the observed co-occurrence patterns among linguistic features.

In fact, there are several unaddressed issues surrounding dimensions identified on functional bases. Although many functional dimensions have been proposed in recent years, few researchers have attempted to relate them to one another or rank them in importance. Consider the following partial list of functional dimensions: informal/formal, restricted/elaborated, contextualized/decontextualized, involved/detached, integrated/fragmented, abstract/concrete, colloquial/literary. Are these all separate dimensions? Do some of them overlap? Are they all equally important? Are they all well-defined in terms of their linguistic

characterization? The approach used here begins to answer these questions. By defining 'dimension' from a strictly linguistic perspective, it is possible to identify the set of dimensions required to account for the linguistic variation within a set of texts. Each dimension comprises an independent group of co-occurring linguistic features, and each co-occurrence pattern can be interpreted in functional terms. The result is an empirical assessment of how many independent dimensions there are; an assessment of which functions are independent and which are associated with the same dimension; and an assessment of the relative importance of different dimensions.

The discussion can be made more concrete by considering some frequency counts in texts 1.1, 1.2, and 1.3. In Table 1.1, I list the frequencies for four linguistic features: passive constructions (including post-nominal modifiers, e.g., *the data [which are] cited*), nominalizations, first and second person pronouns, and contractions. The table includes the raw frequency count and the frequency per 100 words; I use the frequency counts normalized to a text of 100 words to compare the three texts.<sup>1</sup>

The conversational text (1.1) and scientific text (1.2) are quite different with respect to these linguistic features. The scientific text has almost seven passives per 100 words and eleven nominalizations per 100 words; the conversation has no passives and less than one nominalization per 100 words. Assuming that these two texts are representative of their kind, their frequency counts indicate that passives and nominalizations tend to co-occur and thus belong to the same linguistic dimension – when a text has many passives, it also has many nominalizations, as in the scientific text; when a text has few passives, it also has few nominalizations, as in the conversational text. Similarly, these two texts indicate that first and second person pronouns and contractions belong to the same dimension – when a text has many first and second person pronouns, it also has many contractions, as in the conversational text; when a text has few first and

Table 1.1 *Frequency counts for texts 1.1, 1.2, and 1.3 (raw frequency count followed by normalized count per 100 words)*

	passives	nominal-izations	1st & 2nd person pronouns	contrac-tions
conversation	0 / 0	1 / .84	12 / 10.2	6 / 5.1
sci. prose	3 / 6.8	5 / 11.4	0 / 0	0 / 0
panel disc.	2 / 2.2	4 / 4.3	10 / 10.8	3 / 3.2

second person pronouns, it also has few contractions, as in the scientific text.

In addition, we might conclude from these two texts that the passive–nominalization dimension and the pronoun–contraction dimension were in fact parts of the same dimension, because there is a consistent co-occurrence pattern between them. That is, when a text has many passives, it has many nominalizations as well as markedly few pronouns and contractions; conversely, when a text has few passives, it has few nominalizations as well as markedly many pronouns and contractions. For these two texts, knowing the frequency of any single feature allows the researcher to predict the frequencies of the other three features, indicating that they comprise a unified dimension. Passives/nominalizations and pronouns/contractions are not independently related in these two texts – a marked presence of the one set predicts a marked absence of the other. In this sense, dimensions encompass features that consistently occur together and those that consistently complement one another.

Consideration of the panel discussion (text 1.3), however, indicates that passives/nominalizations and pronouns/contractions belong to two separate dimensions. Unlike either the conversation or the scientific text, the panel discussion has high frequencies of all four features. This text confirms the existence of two basic co-occurrence patterns – when a text has many passives, it has many nominalizations; when a text has many first and second person pronouns, it has many contractions. But, the panel discussion shows that these two co-occurrence patterns do not have a consistent relation to one another. It is possible for a text to have many passives/nominalizations and few pronouns/contractions (e.g., the scientific text); it is possible to have many pronouns/contractions and few

<sup>1</sup> Raw frequency counts cannot be used for comparison across texts because they are not all the same length. That is, long texts will tend to have higher frequencies simply because there is more opportunity for a feature to occur; in these cases, the higher count does not indicate a more frequent use of the feature. Comparing the frequency per 100 words eliminates this bias. These normalized frequencies are computed as follows:

$$(\text{actual frequency count} \div \text{total words in text}) \times 100$$

For example, the normalized frequency of contractions in text 1.1 is:

$$(6 \div 118) \times 100 = 5.1$$

passives/nominalizations (the conversation); and it is also possible to have many occurrences of both sets of features (the panel discussion). In fact, it is possible for a text to have few passives/nominalizations and few pronouns/contractions, as the following text sample from a novel shows:

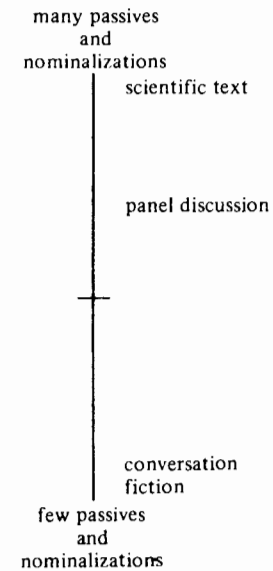
*Text 1.4: Fiction – K 4*

***She became aware that the pace was slackening; now the coach stopped. The moment had come. Upon the ensuing interview the future would depend. Outwardly she was calm, but her heart was beating fast, and the palms of her hands were damp.***

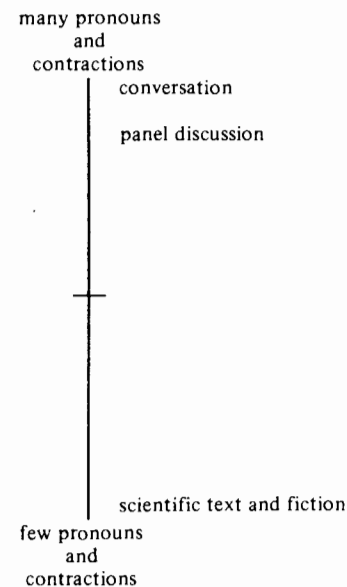
This text has no passives, no nominalizations, no first or second person pronouns, and no contractions. While this distribution further confirms the two basic co-occurrence patterns identified above – passives co-occurring with nominalizations and pronouns co-occurring with contractions – it also confirms the conclusion that these two patterns belong to two independent dimensions. These two dimensions can be plotted to illustrate their independent status, as in Figures 1.1, 1.2, and 1.3. Figure 1.1 shows that conversation and fiction are alike with respect to the passive/nominalization dimension, as are the scientific text and panel discussion. Figure 1.2 shows a different pattern for the dimension comprising first and second person pronouns and contractions: the conversation and panel discussion are alike, as are the scientific and fictional text. The pattern defined by these two dimensions together is shown in Figure 1.3.

Other linguistic dimensions comprise different sets of co-occurring features. For example, in the above four text samples past tense verbs and third person personal pronouns seem to represent a third co-occurrence pattern. Table 1.2 shows that the scientific text has no past tense verbs and no third person pronouns, that the conversation and panel discussion have a few past tense verbs and no third person pronouns, and that the fiction text has a very frequent number of both past tense verbs and third person pronouns. This co-occurrence pattern is independent from the above two patterns, as shown by Figure 1.4.

Once the linguistic co-occurrence patterns are identified, the resulting dimensions can be interpreted in functional terms. The co-occurrence patterns by themselves are not very interesting. Instead, we want to know why these particular sets of features co-occur in texts; we want to know



**Figure 1.1** One-dimensional plot of four genres: nominalizations and passives



**Figure 1.2** One-dimensional plot of four genres: first and second person pronouns and contractions

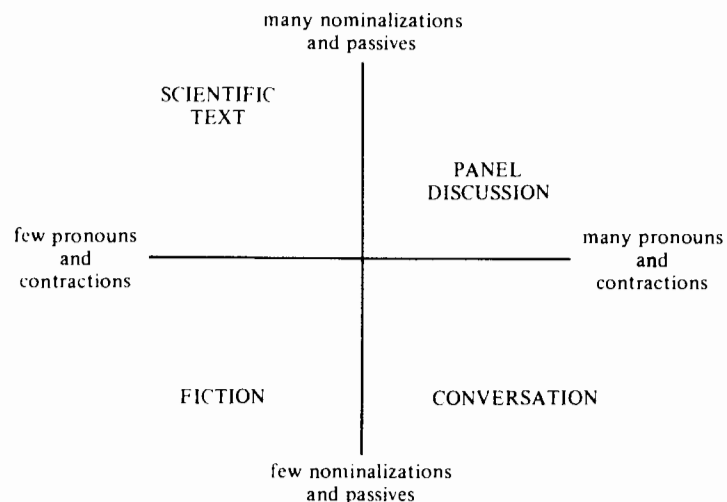


Figure 1.3 Two-dimensional plot of four genres

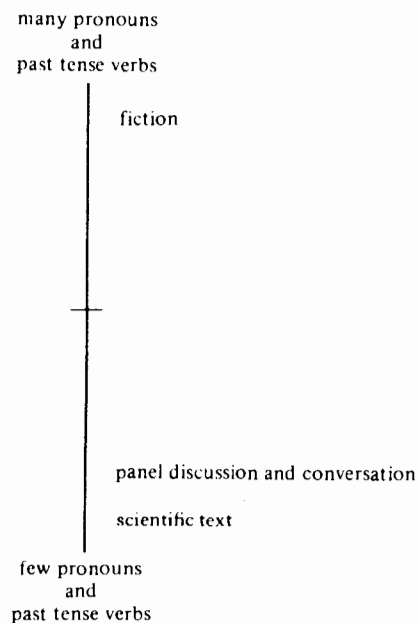


Figure 1.4 One-dimensional plot of four genres: third person pronouns and past tense verbs

Table 1.2 Additional frequency counts for texts 1.1, 1.2, 1.3, and 1.4 (raw frequency count followed by normalized count per 100 words)

	past tense	3rd person pronouns
conv.	2 / 2.1	0 / 0
sci. prose	0 / 0	0 / 0
panel disc.	3 / 2.5	0 / 0
fiction	6 / 14.3	4 / 9.5

what functional or situational parameters relate to the co-occurring sets of features, influencing their systematic use across a range of texts. For example, from a functional perspective, contractions and first and second person pronouns share a colloquial, informal flavor. They are used in interactive situations that require or permit rapid language production. In the present case, they are used frequently in the conversation and the panel discussion, which are both interactive situations. The linguistic dimension of first and second person pronouns and contractions might thus be interpreted as the surface manifestation of an underlying interactive functional dimension.

The other dimensions could be interpreted through a similar process. The co-occurrence pattern between passives and nominalizations can be interpreted as representing an underlying abstract or informational focus. The co-occurrence pattern between past tense verbs and third person pronouns can be interpreted as representing an underlying narrative focus. Any interpretations need to be verified and refined by analysis of the co-occurring features in particular texts. Through this approach, though, we can proceed from the linguistic features that are in fact used systematically in texts to an account of the underlying functional dimensions of English. In this way, we can identify the functional dimensions that are important enough to be systematically marked, and we will be able to specify the extent to which different discourse functions are independent or overlapping.

Once linguistic dimensions are identified and interpreted, they can be used to specify the 'textual relations' among different kinds of texts in English. Each text can be given a precise quantitative characterization with respect to each dimension, in terms of the frequencies of the co-

occurring features that constitute the dimension. This characterization enables a direct comparison of any two texts with respect to each dimension. The textual relations between two texts are defined by a simultaneous comparison of the texts with respect to all dimensions.

Comparison of texts with respect to any single dimension gives an incomplete, and sometimes misleading, picture. For example, consider texts 1.1–1.4 again. If we considered only the passive/nominalization dimension, we would conclude that the fiction text and conversation are linguistically similar and that the scientific text and panel discussion are similar; and that the first two are quite different from the second two. If we considered only the first and second person pronouns/contractions dimension, we would arrive at a quite different set of conclusions: that conversations and panel discussions are quite similar, fiction and academic prose are quite similar, and these two sets of texts are quite different from each other. Finally, considering only the past tense/third person pronouns dimension would lead us to conclude that fiction is very different from the three other texts, which are in turn quite similar to one another. All of these conclusions regarding similarities and differences among texts are inadequate, because the relations among texts cannot be defined unidimensionally. Fiction is not simply similar to or different from scientific prose; rather, it is more or less similar or different with respect to each dimension. Given that the linguistic variation among texts comprises several dimensions, it is no surprise that the relations among texts must be conceptualized in terms of a multi-dimensional space.

The example discussed in this section is extremely simplistic and intended to be illustrative only. To uncover the strong co-occurrence patterns that actually define linguistic dimensions in English, we need to analyze much longer texts, a much larger number of texts taken from many genres, and frequency counts of many linguistic features. Those features that co-occur in different texts across several genres are the ones that define the basic linguistic dimensions of English. A representative selection of texts and linguistic features for analysis is thus a crucial prerequisite to this type of analysis; the range of possible variation must be represented in the texts chosen for analysis, and the range of possible co-occurrence patterns must be represented in the features chosen for analysis. These prerequisites are discussed fully in Chapter 4.

### 1.3 Theoretical bases for the notion of 'dimension'

The notion that linguistic variation must be analyzed in terms of sets of co-occurring features has been proposed in several places. Ervin-Tripp (1972) and Hymes (1974) discuss co-occurrence relations among linguistic features in terms of 'speech styles,' a variety or register that is characterized by a set of co-occurring linguistic features. Brown and Fraser (1979:38–9) emphasize that:

it is often difficult, or indeed misleading, to concentrate on specific, isolated [linguistic] markers without taking into account systematic variations which involve the cooccurrence of sets of markers. A reasonable assumption is that socially significant linguistic variations normally occur as varieties or styles, not as individual markers, and it is on those varieties that we should focus.

Although the theoretical importance of co-occurrence patterns among linguistic features has been well established by these researchers, the empirical identification of salient co-occurrence patterns in English discourse has proven to be difficult. One of the few studies to propose specific sets of co-occurring features is Chafe (1982). This study focuses on two fundamental differences between typical speaking and writing – that speaking is faster than writing, and that speakers interact with their audiences to a greater extent than writers – and it proposes an underlying dimension associated with each of these situational differences: integration/fragmentation and detachment/involvement. Along the integration/fragmentation dimension, integration is marked by features that function to pack information into a text, such as nominalizations, participles, attributive adjectives, and series of prepositional phrases; fragmentation is marked by clauses in succession without connectives or joined by coordinating conjunctions. Along the detachment/involvement dimension, detachment is marked by passives and nominalizations; involvement by first person pronouns, emphatic particles, and hedges. Chafe describes conversational texts as fragmented and involved, showing that they have many loosely joined clauses and many involved features such as first person pronouns and emphatics. He describes academic texts as integrated and detached, showing that they have many features like participles, attributive adjectives, nominalizations, and passives. This study is exemplary in that it recognizes the need to discuss linguistic variation among texts in terms of co-occurring features and actually to identify two dimensions of such variation.

Several other researchers have looked at the distribution of linguistic

## 8 Extending the description: variation within genres

### 8.1 Genres and text types

Genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form. It is also possible to consider groupings of texts that are derived on the basis of linguistic form. In other work (Biber forthcoming) I distinguish 'genres' from 'text types': genres characterize texts on the basis of external criteria, while text types represent groupings of texts that are similar in their linguistic form, irrespective of genre. For example, an academic article on Asian history represents formal, academic exposition in terms of the author's purpose, but its linguistic form might be narrative-like and more similar to some types of fiction than to scientific or engineering academic articles. The genre of such a text would be academic exposition, but its text type might be academic narrative.

Genres are not equally coherent in their linguistic characterizations. Some genres have several sub-classes which are quite different from one another; for example, academic prose includes engineering articles, political and historical analyses, and literary discussions. The linguistic form of texts in other genres is simply not highly constrained, and thus these genres permit a relatively wide range of variation; for example, the linguistic characteristics of face-to-face conversation in private academic settings, public social settings, and intimate settings are all different. In an analysis of text types, texts from different genres are grouped together when they are similar in their linguistic form; texts from a single genre might represent several different text types. It is beyond the scope of the present study to identify underlying text types in English; here I consider only the extent to which genre categories are internally coherent and the relations among several sub-genres.

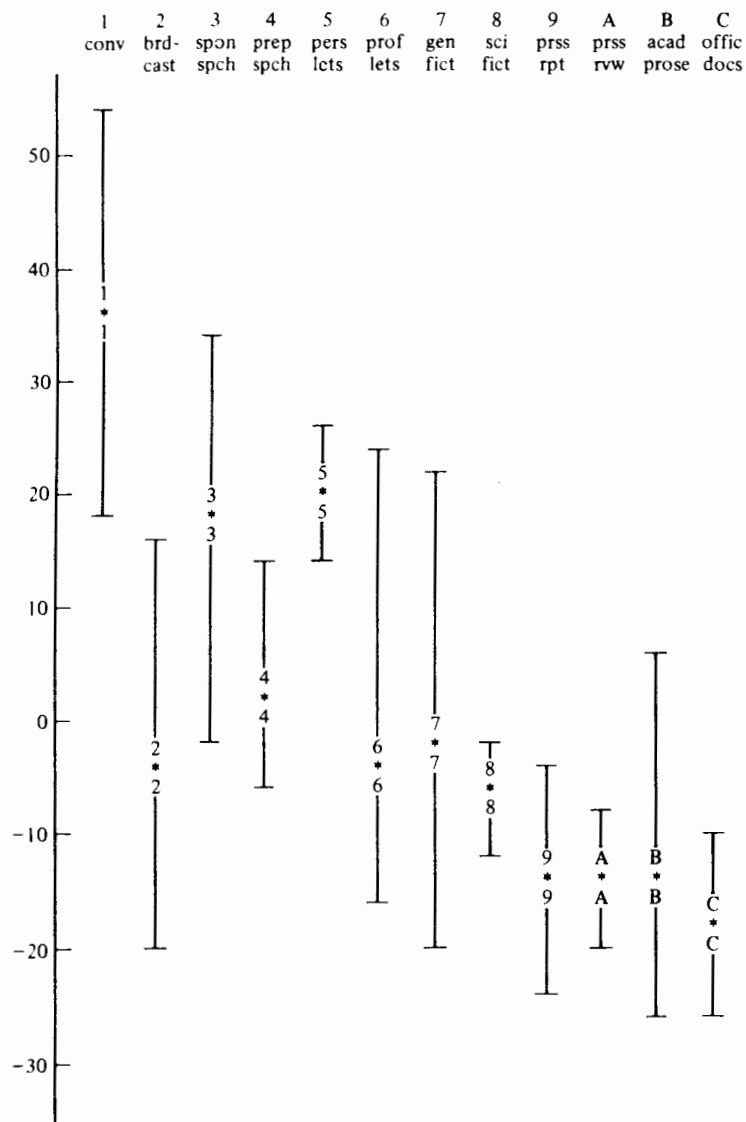
### 8.2 Internal coherence of the genre categories

Figures 8.1–8.6 plot the range of dimension scores found within twelve of the genres used in the present study. These figures plot the maximum, minimum, and mean scores for each genre, taken from Table 7.1. For example, Figure 8.1 plots the range of scores on Dimension 1, 'Informational versus Involved Production'. The first column on this plot represents the range of scores in face-to-face conversation. It shows that the minimum score for a conversational text is around 18, the highest score for a conversational text is around 54, and the mean score for conversational texts is about 35; the actual scores (minimum: 17.7; maximum: 54.1; mean: 35.3) are given on Table 7.1.

A quick look at Figures 8.1–8.6 shows considerable variation in the score ranges. There are much greater ranges on some dimensions than others; for example, there are relatively large ranges on Dimensions 1 and 4, and a relatively small range on Dimension 5. In addition, some genres show much wider ranges than others; compare, for example, academic prose, which has a large range on every dimension, with personal letters, which has relatively small ranges. The range of scores indicates the internal coherence of a genre category – that is, the range of variation possible within a genre.<sup>1</sup> For example, Figures 8.1–8.6 show that academic prose texts can be quite different from one another and still be considered representative of their genre; personal letters are apparently much more similar to one another in their linguistic form.

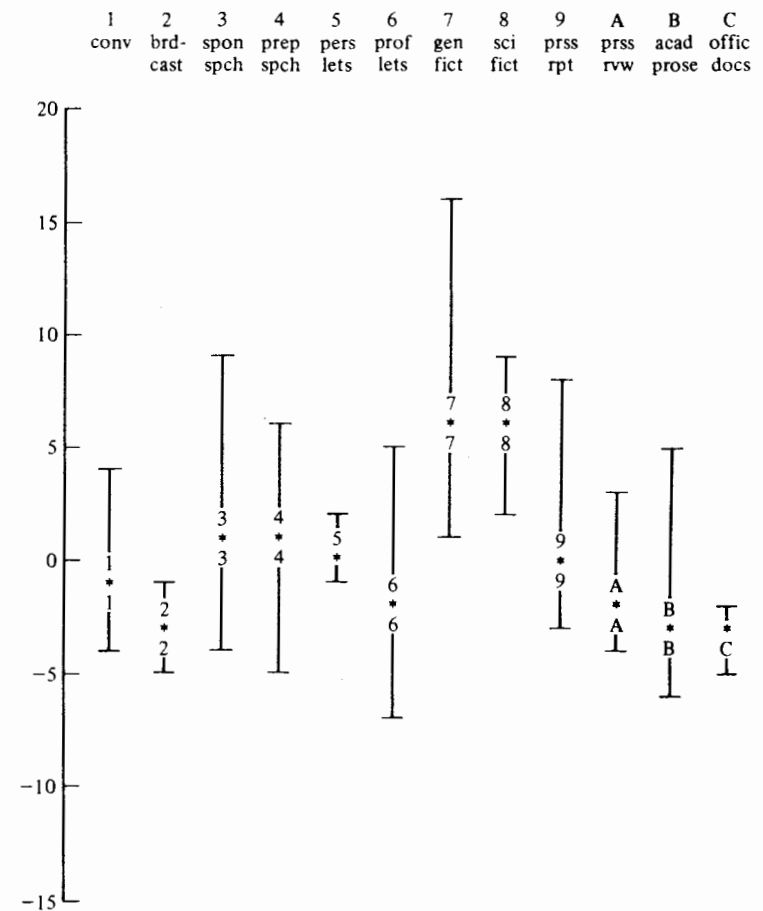
Differences within genres can be considered from two perspectives. First, some of the genres used here include several well-defined sub-genres, and the variation within the genre is due in part to variation among the sub-genres. For example, in the LOB corpus, academic prose is divided into seven sub-categories: natural science, medical, mathematics, social science, politics/education, humanities, and technology/engineering. Due to the differences among these sub-genres, the dimension scores for academic prose have quite large ranges. Some other genres, however, are simply not well-constrained or defined. For example, conversation shows large ranges on most of the dimensions, even though there are no clear-cut sub-genre distinctions within conversation.

<sup>1</sup> Maximum and minimum scores plot the total range of variation within a genre; plots of the standard deviations would indicate how tightly the scores within a genre are grouped around the mean score; see Table 7.1 and the discussion in Section 7.1.

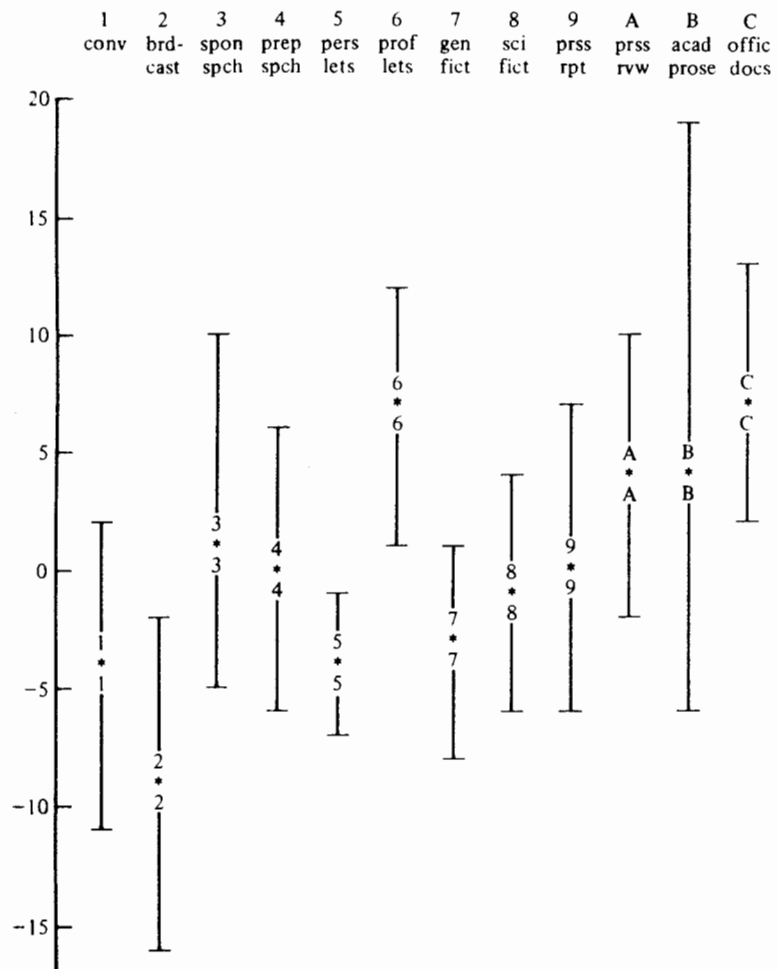


**Figure 8.1** Spread of scores along Dimension 1 ('Informational versus Involved Production') for selected genres (\* marks the mean score of each genre)

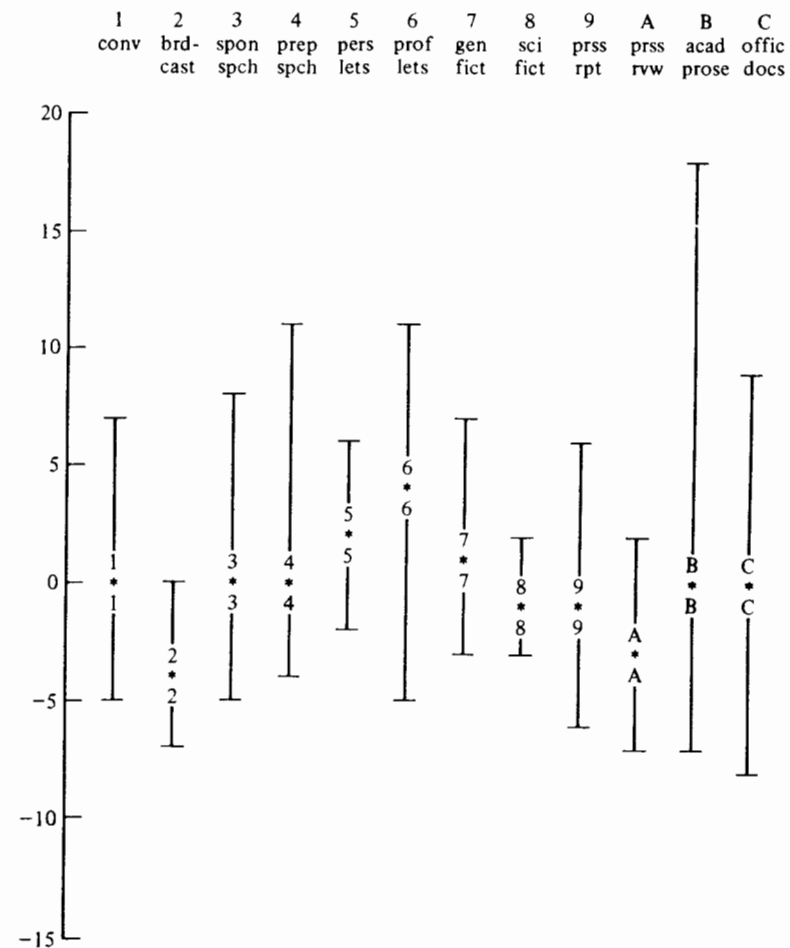
(Key: conv = face-to-face conversation; brdcast = broadcasts; spon spch = spontaneous speeches; prep spch = prepared speeches; pers lets = personal letters; prof lets = professional letters; gen fict = general fiction; sci fict = science fiction; prss rpt = press reportage; prss rvw = press reviews; acad prose = academic prose; offic docs = official documents.)



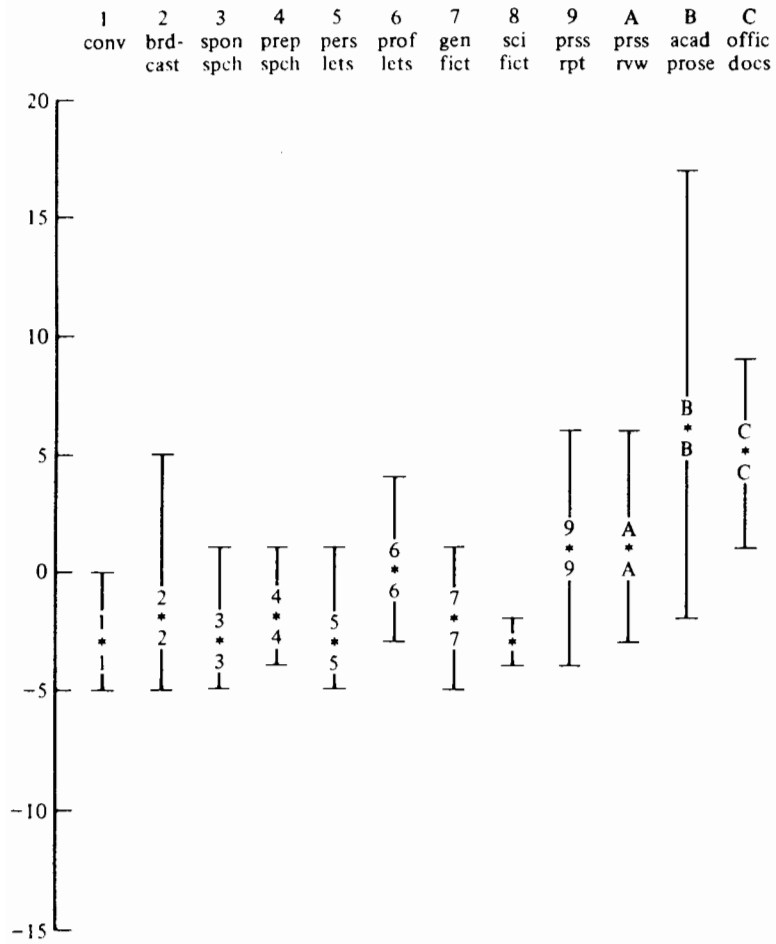
**Figure 8.2** Spread of scores along Dimension 2 ('Narrative versus Non-Narrative Concerns') for selected genres (\* marks the mean score of each genre)



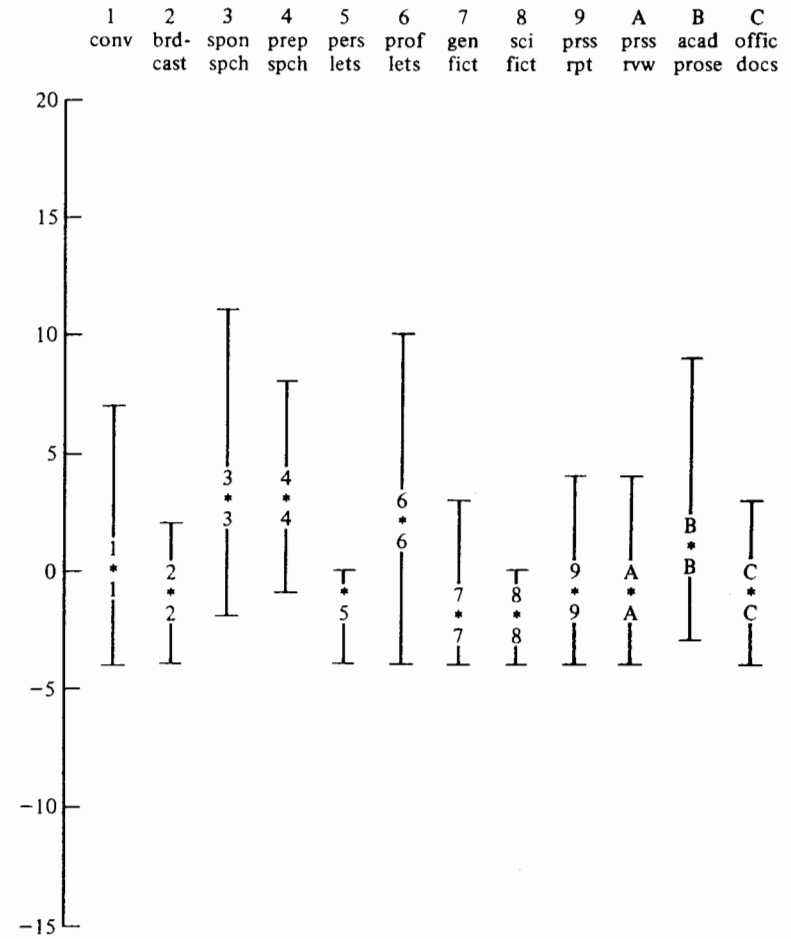
**Figure 8.3** Spread of scores along Dimension 3 ('Explicit versus Situation-Dependent Reference') for selected genres (\* marks the mean score of each genre)



**Figure 8.4** Spread of scores along Dimension 4 ('Overt Expression of Persuasion') for selected genres (\* marks the mean score of each genre)



**Figure 8.5** Spread of scores along Dimension 5 ('Abstract versus Non-Abstract Information') for selected genres (\* marks the mean score of each genre)



**Figure 8.6** Spread of scores along Dimension 6 ('On-Line Informational Elaboration') for selected genres (\* marks the mean score of each genre)

Some of the distributions shown in Figures 8.1–8.6 are surprising. For example, academic prose and official documents are quite different in their internal coherence, although they are quite similar with respect to their mean dimension scores. Official documents actually include several distinguishable discourse types, including government reports, legal documents and treaties, business reports, and a university bulletin; yet this genre is markedly constrained in linguistic form compared to academic prose. The difference between these genres is apparently due to the freedom for personal expression and a personal style in academic prose, whereas official documents are in some sense truly faceless (there is often no acknowledged author) and conform to a much more rigid form. The wide range of scores for academic prose texts is nevertheless surprising and contrary to popular expectation; many studies have considered academic prose to be a stereotypical example of literate discourse, which requires the assumption that academic prose is a well-defined and highly coherent genre.

The fact that there is a wide range of variation within conversation is intuitively sensible, but it is again surprising in relation to the characterization of particular conversational texts as stereotypically oral. The wide range of variation possible within both academic prose and conversation is disconcerting when we consider studies that use a few academic prose texts to represent writing and a few conversational texts to represent speech. I have shown in earlier chapters that no single genre adequately represents writing or speech; the present chapter further shows that no individual text adequately represents academic prose or conversation.

The consistently wide range of variation seen for press reportage in comparison to press reviews is apparently due to the range of sub-genres within press reportage (politics, sports, society, spot news, finance, and cultural events). Broadcasts show a wide range of variation on several dimensions because this genre includes coverage of sports events, non-sports events (such as a funeral and a wedding), and scientific demonstrations. On Dimension 2, 'Narrative versus Non-Narrative Concerns', there is very little variation among broadcast texts because they all report events actually in progress.

General fiction shows a considerably greater range of variation than science fiction. In this case, the small sample size for science fiction (only six texts) biases the comparison, since there is less opportunity for variation within that genre. However, the comparison also indicates that

science fiction is more constrained than general fiction; that general fiction apparently deals with a broader range of topics and uses a broader range of styles than science fiction.

Both prepared speeches and spontaneous speeches include political and legal speeches. In addition, prepared speeches include sermons and a university lecture. Despite the greater range of purposes included in the category of prepared speeches, spontaneous speeches consistently show a greater range of scores. This difference might relate to the planning opportunities in the two genres, but it might also simply indicate a greater freedom for personal variation in spontaneous speeches.

Finally, the comparison of personal letters and professional letters is noteworthy: professional letters consistently have a much greater range of scores than personal letters. This is surprising, given the intuitive impression that personal letters are similar to conversation in being relatively unconstrained, and that professional letters are highly constrained. The actual distribution of texts in these two genres indicates that the opposite is the case: personal letters are apparently quite constrained in their linguistic form while professional letters show considerable variation. This generalization is based on only twenty letters, but the difference between these two genres is quite striking. The personal letters studied here have strictly interactional, affective purposes, and they therefore tolerate little variation in linguistic form. Professional letters, on the other hand, have both interactional and informational purposes, and apparently these two concerns can be weighted quite differently from one professional letter to the next, resulting in considerable variation within this genre.

The above characterizations consider the extent of variation possible within particular genres. We can also consider the extent of variation possible with respect to each dimension. Genres show the least internal variation with respect to Dimension 5, indicating that they are relatively uniform in their characterization as abstract or non-abstract. Academic prose is the only genre to show a wide range of variation on this dimension. The range of variation within most genres is also small with respect to Dimension 2, indicating that genres are relatively uniform in their characterization as narrative or non-narrative. Several genres, however, do show considerable ranges on Dimension 2; surprisingly, general fiction shows one of the largest score spreads on this dimension, perhaps due to variation in the amount of dialogue and description in these texts. The other dimensions show greater ranges within the genres.

On Dimension 1, only personal letters, science fiction, press reviews, and official documents show relatively small ranges. On Dimension 4, only broadcasts and science fiction show small ranges.

The standard deviations of the genre scores, which are also presented in Table 7.1, provide a different perspective on the coherence of the genre categories. The standard deviation shows how tightly a majority of texts are grouped around the genre mean score. A genre can have a small standard deviation, showing that a majority of texts in the genre are grouped tightly around the mean, yet have a large range, showing that at least some of the texts in the genre are quite different from the mean. This is in fact the case with respect to most of the genres studied here; Table 7.1 shows that most genres have relatively small standard deviations, but Figures 8.1–8.6 show that some texts in many of the genres differ greatly from the mean score, indicating that considerable variation is tolerated within most of these genres.

### 8.3 Relations among sub-genres

It was noted in the last section that the large range of variation within some genres is due to the inclusion of several sub-genres. In the present section, I consider the relations among some of these more specific genre categories.<sup>2</sup> Within the genre 'press reportage', the following sub-types of reportage are considered: political, sports, society, spot news, financial, and cultural; within the genre 'editorials', three sub-types are considered: institutional editorials, personal editorials, and letters to the editor; within the genre 'academic prose', seven sub-genres are considered: natural science, medical, mathematics, social science, politics/education/law, humanities, and technology/engineering; two types of 'broadcasts' are considered: sports and non-sports; and finally, three classes of 'telephone conversations' are considered: personal, between business associates, and between disparates.

Table 8.1 presents descriptive statistics for the dimension scores of each of these sub-genres. Similar to Table 7.1, this table presents the mean score, minimum and maximum scores, range, and standard deviation of each dimension score for each sub-genre. The data presented in this table thus enable comparison of the mean scores for different sub-genres as well as consideration of the internal coherence of the sub-genre

<sup>2</sup> The sub-genres considered here include all of the major sub-category distinctions made in the LOB or London-Lund corpus.

Table 8.1 *Descriptive statistics for specialized sub-genres*

Dimension 1: 'Involved versus Informational Production'  
 Dimension 2: 'Narrative versus Non-Narrative Concerns'  
 Dimension 3: 'Explicit versus Situation-Dependent Reference'  
 Dimension 4: 'Overt Expression of Persuasion'  
 Dimension 5: 'Abstract versus Non-Abstract Information'  
 Dimension 6: 'On-Line Informational Elaboration'

Dimension	Mean	Minimum value	Maximum value	Range	Standard deviation
----- Political Press Reportage -----					
Dimension 1	-17.1	-22.6	-11.9	10.7	3.2
Dimension 2	0.8	-2.6	2.6	5.2	1.5
Dimension 3	-0.9	-6.2	5.6	11.7	3.5
Dimension 4	0.6	-3.4	3.4	6.7	2.0
Dimension 5	0.6	-1.6	2.8	4.5	1.7
Dimension 6	0.4	-2.8	3.9	6.7	1.9
----- Sports Press Reportage -----					
Dimension 1	-14.7	-22.7	-10.2	12.4	4.1
Dimension 2	-0.4	-2.0	1.1	3.0	1.2
Dimension 3	-1.2	-4.1	1.6	5.7	2.1
Dimension 4	-0.5	-4.9	2.5	7.4	2.7
Dimension 5	0.1	-3.3	3.6	6.9	2.2
Dimension 6	-1.5	-3.1	0.8	3.9	1.4
----- Society Press Reportage -----					
Dimension 1	-16.1	-22.4	-8.3	14.1	7.2
Dimension 2	-0.4	-2.5	1.0	3.5	1.9
Dimension 3	1.0	-2.1	6.5	8.6	4.8
Dimension 4	-2.1	-4.6	1.3	5.9	3.1
Dimension 5	-0.9	-1.6	0.3	2.0	1.1
Dimension 6	-2.1	-4.0	0.8	4.9	2.6
----- Spot News Reportage -----					
Dimension 1	-13.9	-18.7	-7.6	11.1	3.7
Dimension 2	2.1	-0.4	7.7	8.1	2.5
Dimension 3	0.4	-3.7	4.4	8.0	3.0
Dimension 4	-1.2	-3.7	5.7	9.4	2.9
Dimension 5	1.6	-1.4	5.0	6.5	2.4
Dimension 6	-1.6	-3.6	0.6	4.2	1.4
----- Financial Press Reportage -----					
Dimension 1	-17.6	-24.1	-12.4	11.7	4.9
Dimension 2	-2.0	-2.7	-1.3	1.3	0.5
Dimension 3	-0.2	-1.7	1.7	3.3	1.7
Dimension 4	-1.1	-6.0	3.4	9.5	4.0
Dimension 5	2.7	-1.5	5.5	7.0	3.1
Dimension 6	-0.8	-3.2	1.1	4.4	2.2