SPOKEN AND WRITTEN TEXTUAL DIMENSIONS IN ENGLISH: RESOLVING THE CONTRADICTORY FINDINGS

DOUGLAS BIBER

University of Southern California

Although similarities and differences between speech and writing have often been studied, contradictory claims concerning the linguistic relationship between the two modes are still common. These contradictions can arise from basing global conclusions on restricted methodologies—such as assigning undue weight to individual linguistic features, or to choice of particular text samples and text types. The present study uses a 'multi-feature/multi-dimension' approach, which includes a broad range of linguistic features and text types in a single quantitative analysis, to provide a global description of similarities and differences among spoken/written text types in English. The distribution of 41 linguistic features in 545 text samples of approximately 2000 words each is subjected to factor analysis (a multivariate statistical technique). Three underlying textual dimensions are identified: Interactive vs. Edited Text, Abstract vs. Situated Content, and Reported vs. Immediate Style. To demonstrate the value of the multi-feature/multi-dimension approach, the specific findings of earlier studies are reconciled within the model proposed here.*

INTRODUCTION

1. The study of communicative competence (e.g. Hymes 1972) has provoked an interest in discourse analysis, including the study of linguistic usage in both speech and writing. Sociolinguistic research—challenging the claim that speech is unsystematic, and does not represent the true linguistic structure of a language—has accepted both spoken/written texts for analysis, necessitating an evaluation of the relationship between the two modes. That is, given that both speech and writing are legitimate forms of language, it is important to know in what ways and to what extent they are the same or different.

Various linguistic differences between speech and writing have been proposed. Writing is typically claimed to be more complex, elaborate, and explicit (O'Donnell 1974, Kroll 1977, Olson 1977), more decontextualized (Kay 1977) or detached (Chafe 1982), less personally involved (Blankenship 1974, Chafe & Danielewicz 1985), characterized by a higher concentration of new information (Brown & Yule 1983, Kroch & Hindle 1982), and more highly organized (Akinnaso 1982, Gumperz et al. 1984). Linguistic differences between speech and writing have been attributed to differing processing constraints (Kroch & Hindle) and to differing situational characteristics (Rubin 1980, Olson & Torrance 1981). Chafe & Danielewicz explicitly distinguish these two factors, claiming that features like degree of lexical precision are determined by processing considerations (the restrictions of real-time production in speech vs. opportunity for extensive editing in writing), while features like 1st person pronouns (indicating degree of 'involvement' between speaker/writer and the audience) are conditioned by situational considerations.

* I would like to express thanks to Niko Besnier, Pat Clancy, Bill Grabe, and Deborah Tannen for useful criticisms of earlier versions of this paper; and to Ed Finegan for suggestions and criticisms throughout the research project reported here, in addition to comments on the present paper.

In earlier studies, speech and writing were considered as coherent wholes, and findings were typically presented as general linguistic characteristics of a single dimension in which the two modes were distinguished. A central finding of the present study, however, is that no single dimension adequately accounts for the range of linguistic variation across spoken/written texts. The communicative possibilities offered by a language are complex, and there is no reason to expect a single dimension to be the central discriminator among all text types. Work by Hymes, Gumperz, Labov, and others has described systematic linguistic variation across numerous social and situational dimensions: social class, ethnic group, formality, and intimacy, to name just a few. Thus the claim that the relations among a broad range of text types will be multi-dimensional fits our general knowledge of language use in society, and forms the basis of the present study.

Several researchers have recently noted that few (if any) absolute differences exist between speech and writing: register, topic, and other communicative factors are important in the observed differences among spoken/written text types (see Chafe & Danielewicz 1985, Beaman 1984, Kroch & Hindle 1982, Tannen 1982a). Chafe 1982 suggests that different linguistic features can serve the same communicative function in texts, and proposes two underlying DIMENSIONS—the underlying parameters of linguistic variation among texts: 'integration/fragmentation' and 'detachment/involvement' (cf. Tannen 1982a,b, Chafe & Danielewicz). Chafe's intuitive groupings, however, require confirmation (see Redeker 1984).

A primary need at present is further empirical research to identify the set of underlying textual dimensions which define similarities and differences among text types in English. The relations among multiple linguistic features in multiple texts are too complex to be tracked by simple inspection. Multivariate statistical techniques, such as factor analysis, are required for this task. The present study uses factor analysis for empirical identification of groups of linguistic features which co-occur with a high frequency in texts, indicating a communicative function shared by the features. These groupings represent underlying textual dimensions, which allow us to specify the nature and extent of the similarities and differences among text types. The resulting analysis serves as a guide to further research, and provides an explanatory account of previous research on speech and writing.

1.1. CONTRADICTORY FINDINGS IN PREVIOUS RESEARCH. Most previous linguistic studies of speech and writing have compared frequency counts for certain linguistic features across a few spoken/written text types (e.g. face-to-face conversation, academic prose). A wide range of claims has been presented on the basis of these frequency counts: e.g., Poole & Field 1976 and Halliday 1979 claim that speech is structurally more complex; Blankenship 1962 concludes that there are no important linguistic differences between the two; but most researchers (e.g. O'Donnell 1974, Chafe 1982, Olson 1977) find writing to be more complex, explicit, and integrated.

In a few cases, the cause of these apparent contradictions is simply definitional. Thus Blankenship 1962, 1974 finds sentence length in speech and writing

to be nearly the same; other researchers (e.g. O'Donnell and Poole & Field) find the mean length of sentences, clauses, or T-units (i.e. units consisting of a main clause plus all associated dependent clauses) to be considerably longer in writing. A major problem here is that it is difficult to define 'sentence' in English speech; and since most studies do not define their particular uses of the term, we have no basis for comparison.

Other discrepancies are not explained so easily. One contradiction concerns the extent of subordination. Most studies have found that writing has a much higher degree of subordination and other markers of 'structural complexity' (O'Donnell, Kroll, Kay, Chafe, and Brown & Yule). Other studies contradict these results, finding little difference between the over-all number of subordinate clauses in speech and writing (Poole & Field 1976, Cayer & Sacks 1979). In addition, Poole & Field find a higher index of embedding in speech; and Halliday finds more 'complex sentence structures' (i.e. more subordinate clauses) in speech. Finally, Beaman, in a more careful analysis of subordination, notes that previous studies have been overly simplistic: different subordinate constructions may have differing communicative functions.

A second contradiction involves the frequency of passive constructions—which are associated with decontextualization or detachment, and therefore claimed to be characteristic of writing (e.g. Chafe and Brown & Yule). In contrast to most researchers, Blankenship 1962 finds only slightly more passives in writing. Poole & Field find few passives in either mode.

Additional contradictions could be cited—e.g. concerning the use of foregrounding and backgrounding devices to mark informational relations, or extent of elaboration as shown by the distribution of adjectives and adverbs. It is evident that the relationship between speech and writing is highly complex; however, the methodologies of most quantitative studies have been restricted, and have not provided an adequate basis for global conclusions concerning this relationship. The restrictions include the following:

(a) Assigning undue weight to particular linguistic features: Most quantitative studies have used between three and ten features, and have considered a differential distribution in any of them to be important evidence.

(b) Assigning undue weight to individual texts: Most studies have been based on few texts, so that idiosyncratic differences can have a major influence.

(c) Assigning undue weight to the text types chosen for analysis: Most studies have compared two text types, one spoken (e.g. conversation) and one written (e.g. academic prose), and have not controlled for the different communicative tasks represented by those text types.

In spite of such restrictions, earlier studies have in fact attempted to reach global conclusions—which have thus proved contradictory. A new approach is needed, combining a much broader perspective with a detailed empirical analysis of the linguistic features involved.

1.2. A 'MULTI-FEATURE/MULTI-DIMENSION' APPROACH. The present study provides a framework that avoids the restrictions listed above, while taking advantage of the strengths of quantitative analysis. This is done through a

'multi-feature/multi-dimension' approach, using large-scale text corpora and multivariate statistical techniques (see Biber 1985). First, the large-scale corpora provide a data base of several hundred text samples, representing 16 text types (e.g. conversation, broadcasts, planned speeches, academic prose, fiction). These cover a broad range of variation—associated with differences in situation, purpose, and processing/production strategies—in both speech and writing.

Second, the present study analyses the distribution of 41 lexical and syntactic features, representing a broad range of communicative functions such as content elaboration, or interaction of communicative participants. Factor analysis is used to group the linguistic features empirically into clusters that co-occur with a high frequency in texts. Its use is based on the proposal that such frequent groupings of features in texts indicate a shared, underlying, communicative function. Analysis of the processing and situational differences which characterize types of spoken/written discourse helps interpret the textual dimensions underlying these groupings. In addition, the results of earlier studies are central to the interpretation process, since the identification of a communicative function which underlies a grouping depends on knowledge of the communicative functions of individual features. In a final step, inspection of the distribution of these clusters of features among the 16 text types helps to interpret the nature of the underlying dimension further.

In contrast to the assessment of Akinnaso, the present study demonstrates that a quantitative approach need not be narrow or theoretically uninteresting.[1] It ultimately identifies three separate textual dimensions which describe the similarities and differences among spoken/written text types. These dimensions provide a unified linguistic analysis of a broad range of speech and writing that can be used to explain the contradictory findings of earlier research.

Preliminaries

2.1. Linguistic features used in the comparison. Prior to any comparison of text types, a principled decision must be made concerning the linguistic features to be used. For the purposes of this study, previous research was surveyed to identify potentially important linguistic features and textual dimensions. No a-priori commitment to the importance of an individual linguistic feature or the validity of a textual dimension was made; rather, a wide range of POTENTIALLY important linguistic features was taken from dimensions that might turn out to be valid. In a subsequent step, the dimensions actually accounting for systematic textual variation were identified empirically through a factor analysis.

A survey of previous research on spoken/written differences identified 41

[1] A few quantitative studies using multivariate statistical techniques to analyse text variation have appeared previously. Carroll 1960, Marckworth & Baker 1974, and Horvath 1985 (Ch. 9) use factor analysis, discriminant analysis, and principal-components analysis, respectively, to study aspects of variation among texts. Dubuisson et al. 1983 use multi-dimensional scaling to analyse the development of syntactic complexity in three spoken and written text types. Finally, Poole 1973 uses factor analysis to investigate linguistic code differences from a discourse perspective.

potentially important linguistic features, listed in the Appendix, below. The features are organized in terms of eight communicative functions most commonly associated with each feature. The hypothesized distinctions are:

(a) Writing has a more detached style—shown, for example, by the frequency of passives and nominalizations.

(b) Writing has a more elaborated style, as in the use of subordinate clauses and prepositional phrases.

(c) Writing has a more explicit level of expression, as shown by type/token ratio or precise vocabulary.

(d) Writing has a more explicit marking of informational relations, e.g. cleft constructions.

(e) Speech uses a more inexplicit, informal style of expression, as shown by informal vocabulary items or the general-reference pronoun *it*.

(f) Speech refers more to interactional features, e.g. by using 1st and 2nd person pronouns or questions.

(g) Speech is more situated in a physical/temporal context, as evidenced by place and time adverbs.

(h) Speech and writing differ in their use of verb tenses and aspects, e.g. the past and present tenses.

The factor analysis discussed in §3 shows that some of these features function as hypothesized in previous studies; but that other features must be re-analysed in terms of dimensions, and still others show no systematic co-occurrence patterns. It is necessary, however, to include a wide range of potentially important linguistic features in order to identify the underlying textual dimensions which operate across speech and writing in English.[2]

2.2. THE TEXT SAMPLES used in this study are taken from two large-scale, standard, computerized corpora, which allow significant improvement in the

[2] The counts of these features were done through computer programs written in PL/1, using computerized corpora as input (the untagged version of the LOB Corpus and the London-Lund Corpus—see §2.2, below). Because of the large number of structural options which a particular grammatical construction can take, it was impossible to write programs to count every occurrence. For that reason, the goal of the programs was to capture 70–90% of the occurrences of a construction, with no obvious skewing in one mode or the other. For example, since there is no direct way to identify all verbal forms, I used a list of the 420 most common verbs in the Brown Corpus (Francis & Kučera 1982) to represent all verbs. (This list included all verbs that occurred more than fifty times in the million words of text contained in that corpus.) This decision obviously restricts the completeness of the simple verbal counts—e.g. past and present tense forms—since only occurrences of the items in the list were counted. In addition, this decision influences the count of any construction that is identified through the presence of a verb (e.g. clefts and *that*-clauses). Similar considerations limit the completeness of many other counts. A description of the algorithms used for the counts is available on request. Quirk et al. 1972 was used as the standard syntactic reference during the development of the algorithms.

Some additional linguistic features could have been included in this analysis, but were not—e.g. the frequency of different types of negation (Tottie 1983). Other features were not included because they cannot be analysed automatically—e.g. conjoined phrases and conjoined clauses (cf. Chafe 1982) and features representing different types of cohesion and information structure (Grabe 1984). Future studies should include analysis of both these sets of features.

study of actual language usage. A corpus of this type can provide: (a) a large number of long samples, which are necessary to avoid skewing in the reliability and validity of linguistic counts; (b) text samples taken from several different text types, covering the range of textual variation in the spoken/written domains; and (c) a standardized data base that can be shared among scholars, so that individual studies can be replicated, and results across different studies can be directly compared (see Tottie et al. 1983:7). In addition, computer programs permit an efficient analysis of a large number of linguistic features across a large number of texts. In the present study, 41 linguistic features were counted in 545 text samples, totaling over one million words.

Two separate corpora were used for the text samples. The first is the Lancaster-Oslo-Bergen Corpus of British Written English (known as the LOB Corpus; see Johansson et al. 1978, Johansson 1982); this is drawn exclusively from printed sources published in 1961. It comprises 500 text samples of about 2,000 words each, taken from 15 different genres—e.g. Press (Reportage), Mystery and Detective Fiction, and Learnèd Writings. The total corpus contains approximately a million words of running text.[3]

The second is the London-Lund Corpus of Spoken English (LL; Johansson 1982, Svartvik & Quirk 1980). This is a collection of 87 spoken British English texts of about 5,000 words each. The total corpus contains approximately 500,000 words representing several different speech situations—e.g. Conversation, Broadcasts, and Public speeches.

Sixteen major text types, representing the full range of situational possibilities available in the corpora, were selected for analysis. The distribution of text samples in each text type is given in Table 1, overleaf.[4]

The composition of some of these text types requires elaboration. 'Press reports' include several subclasses: political, sports, society, spot news, financial, and cultural. 'Popular lore' contains texts from popular magazines and books (e.g. *Punch, Woman's Mirror, Wine and Food*). 'Official documents' are primarily governmental, but also include foundation reports, industry reports, and a section from a university catalog. 'Academic prose' combines several subclasses, e.g. natural sciences, medicine, mathematics, social and behavioral sciences, humanities, and technology/engineering. Of the five fictional text types in the LOB corpus, two are included here: general and ro-

[3] The LOB Corpus does not include any examples of written interpersonal communication. For that purpose, I used computerized texts of ten professional letters, generously provided by William Grabe.

[4] Another large collection, the Brown Corpus, contains 500 written texts of American English, but was not used in the present study, because of the possible confounding influence of dialect differences (see Biber 1984). The texts in all three of the large-scale corpora were produced by middle-class, university-educated adults. This coherence in the population under study excludes the possibility of a confounding influence from social differences, but also highlights the need to investigate textual differences across social parameters (see Poole 1979, Kroch & Hindle 1982).

In the present study, the texts from the London-Lund Corpus, which are approximately 5000 words in length, were divided in half to be more closely comparable to the texts in the LOB Corpus (each approximately 2000 words long) and to provide more spoken samples. The frequency counts in all texts were standardized for a text length of 2000 words.

TEXT TYPE	NUMBER OF TEXTS
Written: The LOB corpus (+ professional letters)	
1. Press reports	44
2. Editorial letters	27
3. Skills and hobbies	38
4. Popular lore	44
5. Official documents	30
6. Academic prose	80
7. Belles-lettres	77
8. General fiction	29
9. Romantic fiction	29
10. Professional letters	10
Spoken: The LL corpus	
11. Face-to-face conversation	57
12. Telephone conversation	20
13. Interviews	23
14. Broadcasts	19
15. Spontaneous speeches	9
16. Planned speeches	9
TOTAL	545

TABLE 1. Distribution of text samples.

mantic. 'General fiction' is taken to represent 'serious' work, while 'Romantic fiction' is taken to represent less formal texts (perhaps less carefully written). 'Professional letters' are from one individual to another, but deal primarily with specialized topics. 'Spontaneous speeches' are unprepared public monologs, e.g. from court cases, banquets, or sessions of the House of Commons. 'Planned speeches' are prepared, but without written text; they are taken from sermons, university lectures, court cases etc. Finally, 'Broadcasts' include sports broadcasting and commentary on non-sports events. The other text categories are self-explanatory.

It should be noted that the text types of the corpora are taken as givens in the present study. That is, it is assumed that the distinctions used in the corpora are valid, and that each grouping of texts comprises a single text type. Text types can differ by subject matter, purpose, rhetorical structure, and style—in addition to situational parameters such as the relation between the communicative participants, that of the participants to the external context, and that of the participants to the text itself. At present, the identity of the salient text-type distinctions in English is an unresolved issue (see Faigley & Meyer 1983, Grabe 1984). Additional research is required; however, the distinctions proposed in these corpora have been carefully chosen, and permit the present consideration of relationships among spoken/written text types.

The text types used here represent a broad range of spoken/written situational possibilities (see Rubin 1980, Green & Morgan 1981, Akinnaso 1982 for discussion of the situational differences between speech and writing). Among the written text types, 'Press reports' are directed toward a more general audience than 'Academic prose'; the former involves considerable effort toward maintaining a relationship with its audience, and is concerned with external temporal

and physical situations in addition to abstract information. 'Editorial letters' are less concerned about offending potential readers, but make greater assumptions concerning specific shared background (e.g. particular social issues, or articles previously appearing in the press). 'Professional letters' are structured like academic prose, often stating a thesis with several supporting arguments, but are directed toward a specific audience, and require concern for the interpersonal relationship; they permit a relatively high degree of interaction between participants, and reliance on shared background. Finally, fiction is directed to a very diverse audience, but requires many shared cultural assumptions; it builds its own internal, shared, physical, and temporal context.

Among the spoken text types, public speeches are directed toward broad audiences; they permit little interaction, and a relatively small amount of shared knowledge. (Spontaneous vs. planned speeches differ in the amount of time required or permitted for production, although both allow little time for comprehension, in comparison to written text types.) 'Broadcasts' are directed toward an extremely large audience, but are quite concerned with the temporal and physical contexts being reported. In contrast, 'Interviews' show little concern for the temporal/physical context, but have a high interactional focus—often involving only two direct communicative participants (with a wide audience of more passive participants). Finally, in 'Face-to-face conversation' (and to a lesser extent on the telephone), the interactional focus is primary, usually overshadowing the informational focus. Conversation is characterized by a high degree of interaction and goal negotiability, by considerable effort at maintaining a relationship, and by extensive shared background. Both types of conversation share a temporal context, but the shared physical context is more important in face-to-face conversation.

Other text types not included in the present study differ in further ways; examples include written notes left on the kitchen table, computer conferences, and tape-recorded 'letters'. However, the text types included in this study represent a broad range of situational possibilities across the written/spoken modes.

IDENTIFICATION OF TEXTUAL DIMENSIONS

3.1. OVERVIEW OF FACTOR ANALYSIS. The analysis of speech and writing presented here involves two major steps:

STEP 1: Factor analysis.
- a. Clustering of linguistic features into groups that co-occur with a high frequency in texts.
- b. Interpretation of the factors as textual dimensions, through assessment of the communicative function(s) most widely shared by the features grouped on each factor.

STEP 2: Factor scores as operational representatives of the textual dimensions.
- a. For each factor, computation of a factor score for each text.
- b. Analysis of the distribution of the factor scores among the text types.
- c. Further interpretation of the textual dimensions in light of the relations among the text types with respect to the factor scores.

The aim of Step 1 is to determine which clusters of linguistic features co-occur in the texts and to make initial interpretations of these clusters as textual dimensions. This step assumes that relatively few communicative functions

underlie English discourse; i.e., the 41 linguistic features identified from earlier research are assumed to serve fewer than 41 separate communicative functions. A factor analysis identifies linguistic features that co-occur with a high frequency in texts, and this co-occurrence is taken to indicate a common communicative function shared by these features. Thus each grouping of features, or factor, can be interpreted by consideration of the communicative function most widely shared by the features.

In a factor analysis, a large number of original variables (in this case, the linguistic features) are reduced to a small set of derived variables (the factors). Each factor represents some amount of variation in the original data that can be quantitatively summarized or generalized—a grouping of variables that co-occur with a high frequency in the data. However, only the first few factors are likely to account for non-trivial amounts of the shared variance, and thus be worth further consideration. In the present case, it was determined that five factors account for non-trivial amounts of variance; these were hence retained for further analysis.[5]

Each factor is a simple summation of all the linguistic features, with different features having different weights (known as factor 'loadings'). A restricted set of the linguistic features has salient weights on a given factor; this identifies these features as good representatives of the construct or textual dimension underlying the factor. For example, if the linguistic features in an analysis were past tense, 3rd person pronouns, relative clauses, and infinitives, a factor analysis might produce the following:

FACTOR A = .89 (past) + .61 (3rd pers.) + .10 (inf.) − .19 (rel.cl.)
FACTOR B = −.10 (past) + .29 (3rd pers.) + .56 (inf.) + .65 (rel.cl.)

The number preceding each of the linguistic features is the weight, or factor loading, of that feature for the factor; it indicates the extent to which the feature represents the textual dimension underlying the factor. In the present analysis, features with weights smaller than .35 on a factor are not considered to be salient, and are not included in the interpretation of the factor.[6] Thus, in the above example, past tense (weight .89) and 3rd person pronouns (weight .61)

[5] See Gorsuch 1983, Biber 1985 for a fuller discussion of factor analysis and its application to text analysis. Five factors were retained on the basis of a scree plot of the eigenvalues, which showed a clear break between the fifth and sixth factors. The factors were subsequently rotated using a Promax rotation—a method that permits minor correlations among the factors—since the present application of factor analysis gives no reason to assume orthogonal factors (see Gorsuch, 190 ff.) The intercorrelations among the factors were generally quite small, except for Factors 1–2 (correlation of .58) and Factors 4–5 (correlation of .48). Three linguistic features—style disjuncts, WH-clefts, and split infinitives—did not have salient weights on any of the factors; this shows that they had no systematic distribution with respect to the other features included here. These three features were therefore dropped from the present analysis, and their use for additional studies is in question.

[6] Several methods exist to determine the magnitude of statistically significant loadings in a factor analysis, depending on the number of observations in the analysis (Gorsuch, 208 ff.) Because of the large number of observations in the present analysis, quite small loadings could be considered 'significant'; but an absolute cut-off of .35 was used for the salient loadings.

are the only important features for Factor A; and infinitives (weight .56) and relative clauses (weight .65) are those for Factor B.

Table 2 summarizes the important features for each of the five factors derived through the present analysis. The decimal numbers listed after each linguistic feature show the actual factor loadings of the feature in question.

	FACTOR 1		FACTOR 2		FACTOR 3	
Features with positive weights greater than .35	yes–no questions	.79	nominalizations	.74	past tense	.89
	that-clauses	.76	prepositions	.61	3rd person pronouns	.61
	final prepositions	.68	specific conjuncts	.61	perfect aspect	.47
	pro-verb *do*	.67	agentless passives	.60		
	contractions	.67	*by*-passives	.47		
	I/you	.62	*it*-clefts	.45		
	general hedges	.61	split auxiliary	.42		
	if-clauses	.56	attitudinal disjuncts	.35		
	WH-questions	.52	(word length)	.40		
	pronoun *it*	.49				
	other adverbial subordinators	.48				
	specific emphatics	.46				
	demonstrative BE-WH	.42				
	WH-clauses	.41				
	general emphatics	.41				
	(present tense)	.42				
	(infinitives)	.35				
Features with negative weights greater than .35	word length	−.71	place adverbs	−.57	present tense	−.62
	type/token ratio	−.65	time adverbs	−.55	adjectives	−.40
			relative pronoun deletion	−.50		
			subordinator *that* deletion	−.42		
			(3rd person pronouns)	−.35		

TABLE 2. Summary of the factorial structure of 41 linguistic features. (Features in parentheses are repeated loadings, and are not used in the computation of the factor scores—see §4.)

Features loading on FACTOR 4: relative clauses .65, infinitives .56, (WH-clauses .39), (present tense .38).

Features loading on FACTOR 5: other adverbs .69, specific hedges .39.

Features dropped from the analysis (no salient weights): style disjuncts, WH-clefts, split infinitives.

The negative and positive clusters on a given factor represent two groups of complementary measures.[7] That is, when the features with positive weights occur together frequently in a text, the features with negative weights are markedly less frequent in that text, and vice versa. Taken together, the positive and negative weights represent opposite poles defining an underlying textual dimension. Consider Factor 3: the positive weights are past-time features (past

[7] Positive vs. negative weights on a factor do NOT relate to the importance of those features to the factor. On Factor 3, for example, present tense, with a weight of −.62, is more important than perfect aspect, with a weight of .47.

tense and perfect aspect) and 3rd person pronouns, while the negative weights are present tense verb forms and adjectives. Thus this factor confirms an intuitive judgment that past tense and present tense verb forms are generally complementary. Simply put, a text describing a situation in the past has little need of present tense forms, and vice versa. The positive and negative clusters on the other factors show similar complementary patterns, although the interpretations are not always so intuitively obvious.

3.2. INTERPRETATION OF FACTOR 1. The clusters of linguistic features shown in Table 2 can be interpreted as textual dimensions through an assessment of the communicative function most widely shared by the features. Consider Factor 1: the cluster of features with negative weights includes only word length (i.e. longer words) and type/token ratio (i.e. a more varied vocabulary for a given length of text). The communicative function shared by these features marks a highly exact presentation of informational content in a text, conveying maximum content in the fewest words (cf. Chafe & Danielewicz). A more varied vocabulary reflects extensive use of words having very specific meanings. Longer words also convey more specific, specialized meanings than shorter ones; words become shorter as they are more frequently used and more general in meaning (Zipf 1949). Thus these features share the communicative function of marking highly explicit meaning—a style that tends toward a unique expression for each thought (see Finegan 1985).

The communicative function(s) shared by the positive cluster of features is more complex. Among these features, general hedges and general emphatics, the pronoun *it*, the pro-verb *do*, final prepositions, and contractions can mark a highly inexplicit presentation of meaning—a style in which a single expression can stand for any of several thoughts (cf. Finegan). These features are all reduced forms, tending toward homophonous expression and the loss of lexical meaning. General hedges (e.g. *sort of*) and general emphatics (e.g. *just, really*) are not explicit in their hedging or emphasizing function; e.g., *really* might stand for any of the following more specific emphatics: *absolutely, certainly, extremely, intensely, severely*. *It* stands for an unspecified nominal referent, and *do* stands for an unspecified verbal referent. Contractions result in homophonous forms (e.g. [its] can mean *it is, it has*, or *it*-possessive); and stranded prepositions are indirectly associated with a reduction in meaning—since they usually occur with the caseless relative pronoun *that*, or in relative clauses without a relative pronoun.

In addition, features that can mark a high degree of interpersonal interaction and personal involvement (*I*/*you*, direct yes–no questions, and WH-questions) and four subordination measures (*that*-clauses, *if*-clauses, other adverbial subordinators, and WH-clauses) cluster with the features having positive weights on Factor 1.[8] The occurrence of subordination measures in a pattern which

[8] Specific emphatics and demonstrative-BE-WH constructions also cluster with the features having positive weights on Factor 1, indicating an interactional or involvement function. Thus, although specific emphatics are more carefully chosen than general emphatics, they serve to emphasize the speaker/writer's feelings and attitudes, and so can mark high personal involvement. Demonstrative-BE-WH constructions may also serve an emphatic function, but they require further study.

complements features marking fuller, more explicit informational content disagrees with the conclusions of O'Donnell, Kay, and others—that subordination marks greater elaboration, and is characteristic of writing—but it agrees with the findings of Halliday and of Poole & Field that extensive use of subordination is associated with production constraints characteristic of speech. That is, subordination is apparently an important strategy for expressing fuller content under real-time production constraints, when there is little opportunity to elaborate content through exact lexical choices. Thus Halliday (13) distinguishes between the 'complexities' of speech and writing: spoken language, because it is created and perceived as an on-going process, is characterized by 'an intricacy of movement [and by] complex sentence structures with low lexical density (more clauses, but fewer high-content words per clause)'. In contrast, written language, in which the text is created and perceived as an object, is characterized by 'a denseness of matter [and by] simple sentence structures with high lexical density (more high-content words per clause, but fewer clauses)'. In addition (as shown in §4), subordination of this type can permit a fuller expression of personal involvement and personal feelings (e.g. *I think that ..., I wonder if ..., Let me stress that ...*)

Thus Factor 1 identifies a dimension which characterizes texts produced under conditions of high personal involvement and real-time constraints (marked by low explicitness in the expression of meaning, high subordination, and interactive features)—as opposed to texts produced under conditions permitting considerable editing and high explicitness in the lexical content, but little interaction or personal involvement. The distribution of features shown here suggests the label 'Interactive vs. Edited Text'. This dimension combines both situational and cognitive parameters; in particular, it combines interactional features with those reflecting production constraints (or the lack of them). Given the range of text types included in the present analysis, it is not surprising that these two parameters are combined in a single dimension: highly interactional texts are frequently produced under the greatest production constraints (see §4).

3.3. INTERPRETATION OF FACTOR 2. Here the features with positive weights (nominalizations, prepositions, passives, specific conjuncts, *it*-clefts etc.) share a function which marks a highly abstract, nominal content and a highly learnèd style. Thus the passive often promotes abstract concepts to subject status—while demoting the more concrete agent to object status, or deleting it altogether (cf. Chafe & Danielewicz). Nominalizations compress the information of a clause into a single nominal element (e.g. *destruction of x* corresponding to *Somebody destroys x*); this again results in the loss of an active agentive participant, and the promotion of a more abstract concept. Prepositional phrases figure prominently in this style of discourse, as markers of the varied semantic relations among the entities and concepts in a text. Apparently, as a text becomes more abstract (and semantically complex), and involves a broader range of semantic relations, more prepositional phrases are required to mark these relations. (The exact function of prepositional phrases in relation to the other features on this factor requires further research.) Specific conjuncts and *it*-

clefts also serve to mark the informational relations among different components of a text. Attitudinal disjuncts and adverbs occurring with split auxiliaries (e.g. *He was obviously working hard*) have lower weights on this factor; they apparently serve to mark the author/speaker's attitudes in texts having a highly abstract content. In addition, all the features with positive weights are associated with a high degree of formality and a learnèd style.

In contrast, the features with negative weights on Factor 2 share the marking of very concrete content and more informal style, indicated by high reference to the temporal and physical situation—by means of place and time adverbs—and reduced surface form, through deletion of relative pronouns and subordinator *that*. Place and time adverbs refer directly to an external situation, clearly marking a more concrete, situated content. Deletion of relative pronouns and subordinator *that* mark a reduced correspondence between the surface form and underlying meaning; they are associated with less formal styles, and with speech more than writing (Finegan & Biber 1983). They thus reflect a greater reliance on an external situation than more deliberate styles. Consideration of the features with positive and negative weights suggests the label 'Abstract vs. Situated Content' for the dimension identified by this factor—i.e. a detached, formal style vs. a concrete, colloquial one.[9]

3.4. INTERPRETATION OF FACTORS 3–5. For Factor 3, the features with positive weights (past tense, perfect aspect, and 3rd person pronouns) can all refer to a removed, narrative context; those with negative weights (present tense and adjectives) can be used for more immediate reference. The co-occurrence of adjectives with the present tense apparently indicates the presence of more elaborated content in present-time descriptive or expository texts than in past-time narrative texts; however, this feature needs further study. Over-all, the dimension identified by this factor distinguishes texts with a primary narrative emphasis, marked by considerable reference to a removed situation, from those with non-narrative emphases (descriptive, expository, or other), marked by little reference to a removed situation but by high occurrence of present tense forms. These characteristics suggest the label 'Reported vs. Immediate Style'.

Factors 4–5 are more difficult to interpret than the first three. Factor 4 has only four features with salient weights, and two of these (WH-clauses and present tense) have larger weights on other factors. Factor 5 has only two features with salient weights. Thus neither factor is well represented, and each must be interpreted cautiously. On-going research is considering other measures in relation to these two factors, to test their importance and the validity of the interpretations suggested here.

The communicative function shared by the features on Factor 4 (relative

[9] The present study shows that 'abstract, detached content' is not directly opposed to a high amount of personal involvement, contrary to Chafe's proposal. The groupings of features on Factors 1–2 show that those associated with personal involvement and those associated with high abstraction (or detachment) belong to separate textual dimensions, although the two dimensions are correlated. Professional letters, discussed in §4, illustrate highly abstract texts with a high level of personal involvement.

clauses, infinitives, WH-clauses, and present tense) seems to mark an 'integrative' type of subordination (cf. Chafe), as opposed to that associated with real-time production constraints seen in connection with Factor 1. That is, this type of subordination may be used to package a high amount of information into a text; it is characteristic of 'static' rather than 'dynamic' texts (to use Halliday's terms). If this interpretation is correct, we would expect that features which have been hypothesized as being integrative (e.g. participles) should co-occur with the features of Factor 4, whereas features hypothesized as opposing integration (such as conjoined clauses) should occur in a complementary pattern.

The communicative function shared by the features on Factor 5 (adverbs and specific hedges) seems to mark the author's or speaker's stance in a text. Specific emphatics have a weight of .32 on this factor—too low to be considered salient, but in line with the stated interpretation. Linguistic features which might mark author's stance also occur as parts of Factor 1 (general hedges, general emphatics, and specific emphatics) and of Factor 2 (attitudinal disjuncts and adverbs occurring as split auxiliaries); this indicates that the notion of stance is complex, and requires further research (see Biber & Finegan 1985).

The interpretations of the dimensions underlying these factors are open to refinement, and require further validation. As we learn more about the communicative functions of specific linguistic features, the emphases of some interpretations may shift. The interpretations given above for the last two factors must be considered speculative, since they are not well-represented. For Factors 1–3, however, the groupings of features are quite stable (see the partial replication of this study reported in Biber 1984); thus we can have confidence in the claim that important textual dimensions are being represented here—ones that will be useful for defining relations among spoken/written text types.

A UNIFIED MODEL

4.1. OVERVIEW OF FACTOR SCORES. In §3, we discussed interpretations which result from the factor analysis. In Step 2 (§3.1), derived variables that operationally represent the textual dimension underlying each factor can be computed. These derived variables, known as FACTOR SCORES, allow further interpretation of the textual dimensions by examination of the similarities and differences among the text types with respect to each dimension.

A factor score is computed by summing, for a given text, the number of occurrences of the features having salient weights on that factor. Thus the score for Factor 3 would be computed by adding the number of past tense forms, perfect aspect forms, and 3rd person pronouns (i.e. the features with positive weights), and then subtracting the number of present tense forms and adjectives (i.e. the features with negative weights). For example, one of the fictional texts in this study has 156 past tense forms, 117 3rd person pronouns, 24 perfect aspect forms, 46 present tense forms, and 88 adjectives, resulting in the following factor score for Factor 3:

$$(156 + 117 + 24) - (46 + 88) = 163$$

Some of the linguistic features have salient weights on more than one of the factors (e.g. word length on Factors 1–2); to assure the experimental inde-

pendence of the factor scores, each feature was included in the computation of only one factor score (see Gorsuch, 268). Thus each linguistic feature is included in the score for the factor on which it has the highest weight. Salient weights not used in the computation of the factor scores are given in parentheses in Table 2. Thus, since word length has a weight of −.71 on Factor 1 but only .40 on Factor 2, it is included in the factor score for Factor 1.

The relations among spoken/written text types can be considered through plots of the mean values of the derived factor scores (representing underlying textual dimensions) for each text type (Figures 1–3). That is, a score for each factor was computed for each text (as illustrated above); then the mean of each factor score was computed for each text type. For example, if we had only three fictional texts, with scores for Factor 3 of 163, 187, and 190, then the mean score for fiction on Dimension 3 would be:

$$(163 + 187 + 190) / 3 = 180$$

Figs. 1–3 present the mean scores of each factor score for each text type, showing the relations among the text types along the first three dimensions.

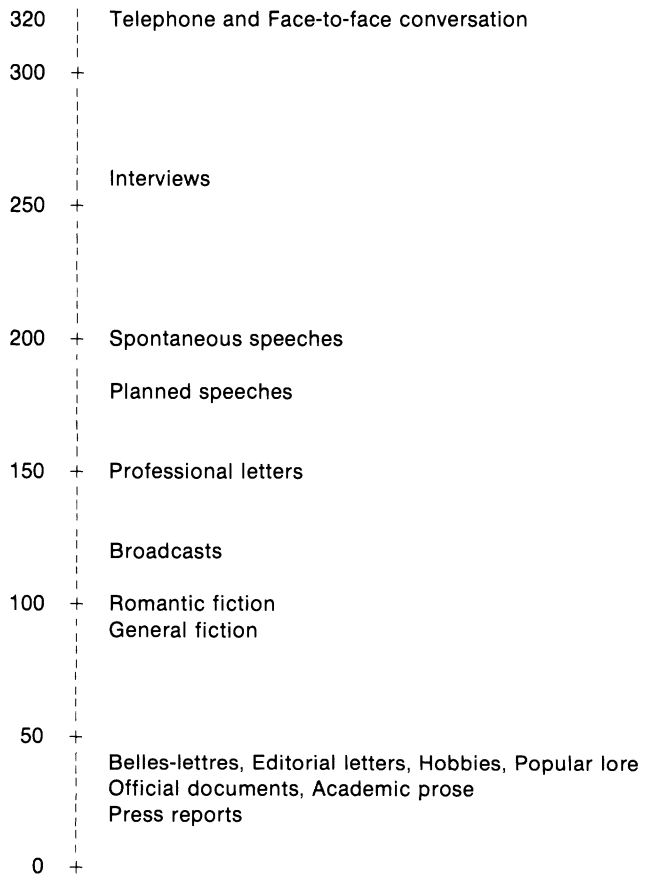

FIGURE 1. Mean scores for Textual Dimension 1: Interactive vs. Edited Text ($F = 257.90$, $p < .0001$, R*R = 88%).

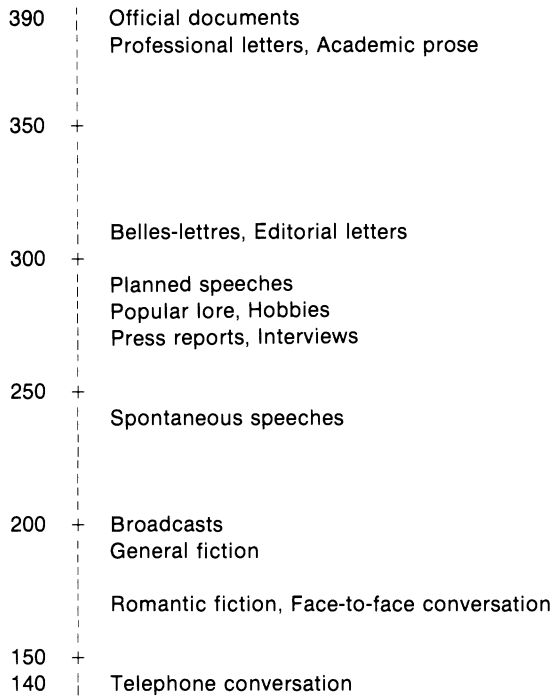

FIGURE 2. Mean scores for Textual Dimension 2: Abstract vs. Situated Content ($F = 82.04$, $p < .0001$, $R*R = 69.9\%$).

Only scores for the first three factors are considered here, since the remaining factors need further research. A statistical procedure called ANOVA shows that there are significant differences among the text types with respect to each of these three factor scores. (See the F and p values reported in Figs. 1–3, where a p value smaller than .05 indicates a statistically significant relationship.) The R*R (squared multiple correlation coefficient) values presented in Figs. 1–3 indicate the importance of each factor score (the percentage of variance in the text type distinctions accounted for by knowing the factor score). Thus Factor Score 1 is the most important: R*R = 88%, meaning that 88% of the variance in the text type distinctions can be accounted for by Factor Score 1. It is followed by Factor Score 2 (R*R = 70%) and Factor Score 3 (R*R = 46%).

4.2. DISTRIBUTION OF TEXT TYPES ALONG TEXTUAL DIMENSION 1. A fuller interpretation of the textual dimensions represented by these three factor scores is made possible by a simultaneous consideration of the distribution of the factor scores across the text types (shown by the plots in Figs. 1–3), the situational or cognitive parameters associated with the distribution, and the linguistic features constituting the factor score. Consider the plot of text types along Dimension 1. The spoken text types tend toward the high end of this dimension, while the written types tend toward the low end. Face-to-face and telephone conversations show the highest values; official documents, academic prose,

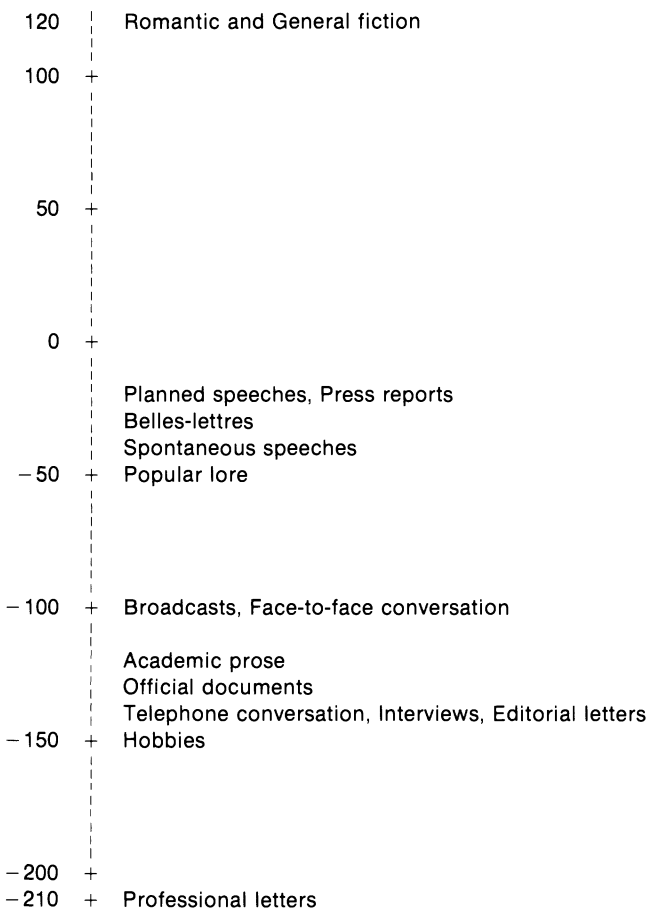

FIGURE 3. Mean scores for Textual Dimension 3: Reported vs. Immediate Style ($F = 29.48$, $p < .0001$, $R*R = 45.5\%$).

and press reports show the lowest. Text types with high scores on this dimension are characterized by frequent occurrences of questions, 1st and 2nd person pronouns, contractions, pro-verb *do*, pronoun *it*, *that*-clauses, *if*-clauses etc. (i.e. the linguistic features with positive weights on Factor 1—see Table 2), plus markedly low values for word length and type/token ratio (the features with negative weights on Factor 1). Types with low scores on this dimension have the opposite characteristics. These characteristics are illustrated by the following two text samples—one from a face-to-face conversation, the other from an official document (a university bulletin).[10]

[10] Text samples are labeled as follows:

CORPUS:TEXT-TYPE.TEXT-NUMBER

For example, Text Sample 1 is labeled LL:1.4, because it is from the London-Lund Corpus, text type 1 (face-to-face conversation), and is text 4 within that type. In the spoken text samples, each line represents an intonation unit.

TEXT SAMPLE 1 (LL:1.4, Face-to-face conversation).

B: it doesn't need to
but it does in fact
by tradition
all the
A: finalists
B: finalists go
and so the others mmm
the others sort of feel
that things won't go on much longer
A: well they really haven't any reason to
because I mean finalists are
B: mmm
A: and they actually do finish
B: exactly
of course they do
A: and the others don't
well
I don't know
B: but I don't think it's feasible
I mean I know this is the first time I've done it
and I'm not in a main line paper
but I'm sure it'll take me all my time to do it
in three weeks
I mean I've seen what it's been like for you
I know ... had more
on the other hand
I must allow myself good time
the first time I do it
A: I don't think I'm going to go on with it
B: are you doing two or one paper this year
A: only one

TEXT SAMPLE 2 (LOB:H.29, Official documents).

The University expects its students to conduct themselves at all times in an orderly manner creditable to the good name of the University. Regulations for the maintenance of good order and discipline are promulgated from time to time ...

The official dates of University terms as published in the Calendar apply to all students. Students (other than new students at the opening of a session and research students) are required to arrive in Hull on the first day of term and, except with the special permission of the Dean of their Faculty, may not go down until the last day. The first and last days of term as published are regarded as travelling days on which no lectures or classes will be held ...

Text Sample 1 illustrates the major linguistic characteristics of texts having high values on Dimension 1: generalized lexical content (*it, sort of, well, really*, pro-verb *do* etc., plus few long words and a restricted vocabulary); a high degree of self-reference and interpersonal interaction (*I, you, are you doing ...?*); and a high degree of subordination, often used to express further details of personal attitudes (*the others ... feel that ..., because, I don't think* [*that*] *it's feasible, I know* [*that*] *this is ..., I've seen what it's been ...*) Text Sample 2 illustrates the opposite characteristics: many long words and a quite varied vocabulary, resulting in an explicit expression of meaning (*an orderly manner creditable to ..., regulations for the maintenance ... are promulgated ...*)—coupled with few generalized content terms, self-reference terms, interactional construc-

tions, or subordinate constructions of the type included in Factor 1 (viz. *that*-clauses, *if*-clauses etc.)

It might be supposed from these examples that Dimension 1 represents a fundamental difference between speech and writing. However, a simple spoken/written distinction cannot explain the over-all distribution of text types shown in Fig. 1. Among the written text types, professional letters show a relatively high score; but among the spoken text types, spontaneous and planned speeches show rather low scores, and broadcasts are very low. Within both speech and writing, the text types which permit the most personal involvement and interaction, and which have the greatest production constraints (imposed by a real-time, on-going interaction), show the highest scores on this dimension;[11] those with the least involvement and the most opportunity for editing score the lowest. Thus the distribution of text types in Fig. 1 shows that this dimension represents the extent of personal involvement, interaction, and opportunity for editing in a text type—a dimension that is independent of physical mode, although there is considerable overlap. This distribution of text types provides support for the proposed label 'Interactive vs. Edited Text'.

It was noted in §3 that this is a hybrid label, reflecting both personal/interactional and cognitive/production parameters. These parameters are conceptually distinct, and could be analysed separately. In fact, consideration of the cognitive/production parameter by itself seems to show an absolute difference between spoken/written text types. Chafe & Danielewicz show that the most carefully planned (but unscripted) speeches have lower type/token ratios than relatively unplanned letters; however, the influence of differences in topic coherence needs further analysis. Finegan & Biber 1984 show an absolute difference between spoken/written text types with respect to those measures of linguistic complexity that reflect the different production strategies of speech and writing (although the relation of personal letters to spoken text types is not considered). In the present study, the combination of production and interactional parameters as part of Dimension 1 indicates the existence of a shared parameter underlying both sets of features. That is, although these two parameters are conceptually distinct, Dimension 1 shows them to be intimately related (at least with respect to the text types used).

4.3. DISTRIBUTION OF TEXT TYPES ALONG TEXTUAL DIMENSION 2. Turning to the dimension of Abstract vs. Situated Content (Fig. 2), we see again that there is considerable overlap with the spoken/written distinction: the written text types (e.g. official documents, academic prose, and professional letters) tend to have high scores, while the spoken text types (e.g. telephone and face-to-face conversations) tend to have low scores. In Table 2, we saw that types with high values on this dimension are characterized by a highly abstract,

[11] High production constraints in conversation are imposed by the on-line nature of communication, with little opportunity for planning or editing. In professional letters (the written text type with the highest score on Dimension 1), the production constraints are self-imposed; i.e., relative to academic prose, fiction, and other highly-edited written types, letters are produced in real time (possibly because of the lesser social value attached to them), and show some of the features indicative of real-time production constraints.

detached content (high frequencies for passives, nominalizations, prepositional phrases etc., coupled with markedly low frequencies for place and time adverbs, relative pronoun deletion, and subordinator *that* deletion); in contrast, types with low scores are characterized by a highly concrete, situated content.

Consider the above text samples again. Sample 2 represents the high pole of this dimension, characterized by features which mark highly abstract content—nominalizations (*regulations, maintenance, permission*); passives (*are promulgated, as published, are regarded, will be held*); and extensive use of prepositional phrases (*for the maintenance of good order, in Hull on the first day of term*). Sample 1 represents the low pole of this dimension; thus it is characterized by few nominalizations, passives, or prepositional phrases, but by frequent references to place and time of interaction (*much longer, in three weeks, this year*) and by reduced surface form (*don't think* [*that*] *it's, I know* [*that*] *this, don't think* [*that*] *I'm*).

As with Dimension 1, a simple spoken/written distinction cannot explain the over-all distribution of text types along Dimension 2. Planned speeches, interviews, and spontaneous speeches all get high scores along this dimension, while general and romantic fiction get some of the lowest. This distribution is in line with the proposed interpretation of Abstract vs. Situated Content, which correlates highly with the deliberate use of a learnèd style. Planned speeches (sermons, university lectures etc.), although containing direct references to the temporal and physical situation of the discourse, are quite abstract in content; thus they show relatively high scores on this dimension. Fiction, although written, does not present an abstract content, and it makes extensive reference to an internally-constructed temporal and physical situation (see Rader 1982); thus it gets a very low score on this dimension. The distribution of text types along this dimension is therefore in agreement with the suggested label 'Abstract vs. Situated Content'.

4.4. DISTRIBUTION OF TEXT TYPES ALONG TEXTUAL DIMENSION 3. The primary distinction along the dimension of Reported vs. Immediate Style (Fig. 3) is fiction, with high mean scores, vs. non-fiction, with low ones (professional letters showing the lowest mean score). The fictional text types are characterized by frequent past tense, perfect aspect, and 3rd person pronoun forms, together with a marked absence of both present tense forms and adjectives (see Table 2). The text types with low scores are marked by the opposite characteristics. Consider Samples 3 and 4:

TEXT SAMPLE 3 (LOB:P.24, Romantic fiction).

> But Mike Deegan was boiling mad now. When the inning was over he cursed the Anniston catcher all the way into the dugout. Phil Rossoff, coming in from left field, stopped at the water fountain for a drink. Mike Deegan was standing beside it, facing the field. He was eyeing the Anniston catcher warming up his pitcher before the inning began. 'Keep your eyes open, sonny,' Mike yelled to the catcher. 'You're in for trouble.'
>
> ... The Anniston manager came right up to the dugout in front of Mike. His face was flushed. 'Deegan,' the manager said, his voice pitched low, quivering. 'That was a rotten thing to do.'
>
> 'For God's sake,' Mike said, waving the manager away. 'Stop it, will you? Tell your guys not to block the plate!'
>
> 'You didn't have to ram him.'
>
> 'That's what you say.'

TEXT SAMPLE 4 (Professional letters; see fn. 3).

> While these initiatives, I concede, border on programmatic activity, it should be stressed that they were taken pursuant to the development of specific policy/programmatic recommendations for consideration by the board of directors in October and the board was informed of them in our report of May. It should also be understood that ... Is the basis of international educational exchange threatened when I hear optimistic forecasts? Perhaps we can agree that foreign students will remain between 2 and 3% of total enrollment. I disagree with your assertion of my claiming that good extended orientation programs help; I doubt that xxxx xxxx has ever viewed his program as orientation and indeed that is part of my disenchantment with xxxx.

Sample 3 shows the typical characteristics of fiction; apart from the dialog sections, it uses only past tense forms, and it has a high frequency of 3rd person pronouns. Sample 4, by contrast, is written consistently in the present tense and contains only one 3rd person pronoun. These distributions reflect the differing purposes of the two text types. Fiction is primarily narrative—a report of past events, combined with fictional dialog. Professional letters are primarily expository and written for a very specific purpose; they are addressed to an individual, and deal with issues immediately at hand.

The other text types range between these two poles. Some, such as planned speeches and press reports, show intermediate mean scores, indicating a mixed style with respect to this dimension: reported past events, combined with informational exposition using present tense forms. The remaining text types, such as conversation and academic prose, have lower mean values on this dimension, indicating a preference for present tense forms, but less consistency in the use of these forms than we find in professional letters. It can be seen from the clustering of text types along this dimension that present tense forms can be used for description of events in progress (as in broadcasting and conversation) or for expository purposes (as in academic prose or professional letters). The distribution of text types along Dimension 3 is thus in agreement with the label 'Reported vs. Immediate Style' suggested in §3.

4.5. THE NEED FOR A MULTI-DIMENSIONAL ACCOUNT OF SPEECH AND WRITING. We have briefly considered the relations among spoken/written text types with respect to each of the three textual dimensions identified by the factor analysis. We have seen that the dimensions are separate: each represents distinct communicative functions, and each identifies a separate set of similarities and differences among text types. If the relations among spoken/written text types were considered in terms of only one dimension, a necessarily incomplete description would result.

Consider the description of professional letters with respect to the three dimensions. Academic prose, official documents, and professional letters are nearly identical with respect to Dimension 2. Thus Sample 4 (from a professional letter) shows a high frequency of nominalizations (*initiatives, development, recommendations, consideration* etc.), of passives (*be stressed, were taken, was informed* etc.), and of prepositional phrases (*to the development of specific ... for consideration by the board ...*) This sample illustrates the highly abstract informational content found in professional letters, making them like academic prose and official documents. But with respect to Dimension 1,

these text types are quite different. Professional letters show high type/token ratio and use of long words, as do academic prose and official documents; but they show considerable personal involvement and interaction with the reader. Sample 4 shows frequent 1st person reference and high use of subordinate clauses to express personal feelings (*should be stressed that ..., should also be understood that ..., I doubt that ...*)—plus frequent interactive features such as 2nd person pronouns and questions (*Is the basis of ...?*) Finally, as shown above, professional letters show the lowest score (or the most 'immediate' style) for Dimension 3. Thus a consideration of any of these dimensions alone would result in an inaccurate description of professional letters. If we considered only the features associated with Dimension 2, we might conclude that professional letters were indistinguishable from academic prose; if we considered only Dimension 1, we might conclude that professional letters were very similar to planned speeches or broadcasts. An adequate description of a text type and its relations to other text types requires a consideration of that text type with respect to all three dimensions.

The need for all three textual dimensions can also be seen from other comparisons. Planned speeches show a pattern similar to that of professional letters, in that both permit considerable personal involvement and interaction—as shown by their relatively high scores on Dimension 1. Both also contain quite abstract content and are not highly situated (this is truer of professional letters than of planned speeches); thus they show high scores on Dimension 2. Sample 5 illustrates these characteristics in a planned speech:

TEXT SAMPLE 5 (LL:12.5, Planned speeches).

does anyone believe
that we would have accepted for the seventies
a degree of freedom of capital movement
that would have aggravated that power of speculative attack on sterling
which we had to fight in the sixties
if Mr. Barber
with inherited
Labour exchange controls
had to admit to a thousand million loss
through a run on sterling
in six days last June
could any Labour government have agreed ...
they can see ten years ahead
but they're not going to tell us
what's going to happen in the next ten weeks
and they still haven't told us ...
well why can't we be told ...
if this statement is rejected
you reject the right of the people to decide
no other resolution
adequately provides
for the British people having the last word
the right of self-determination

Here we see frequent use of the features of interaction and involvement associated with Dimension 1 (*does anyone believe that we ...?; could any ...*

have ...?; *why can't we be told?*)—but also of the passives and nominalizations associated with Dimension 2, which mark an abstract content (*be told, is rejected, resolution, self-determination*). Dimension 2 begins to distinguish between planned speeches and professional letters, showing the speeches to be more highly situated than the letters. For example, consider the direct references to the temporal/physical situation in Sample 5 (*six days last June, ten years ahead*). The difference between these two text types is clearest, however, along Dimension 3. Professional letters show a strong preference for present tense forms and for exclusion of reported events, while planned speeches use both present and past forms. Present tense conveys expository content (e.g., from Sample 5, *if ... is rejected ..., you reject ..., provides*), while past tense reports past events given in support of the expository points (e.g. *would have accepted, we had to fight, had to admit*). Thus, consideration of all three dimensions is required to distinguish these text types.

Broadcasts, as in Sample 6, appear to be quite similar to most written text types with respect to Dimension 1.

TEXT SAMPLE 6 (LL:10.4a, Broadcasts).

and it's Fire Raiser who's coming up to join Carbon
Fire Raiser on the stand side
and right over on the stand side
is Grange Court
coming to the two furlong from home marker
and it's Carbon over on the far side in the sheepskin noseband
in the centre is Fire Raiser in the check cap
and on the stand side is Grange Court and
Carbon's under pressure over on the far side and Fire Raiser's gone to the front
and it's Fire Raiser the leader now ...
and it's Carbon on the far side
Fire Raiser fighting his way back
and it looks as though Carbon is just going to hold him
and at the line
Carbon's the winner

This sample shows few interactional features or indications of personal involvement, although real-time production constraints are quite obvious (shown by the restricted range of vocabulary and high repetition of words). However, a consideration of Dimension 2 shows broadcasts to be quite similar to the majority of spoken text types in being highly concrete and situated. Sample 6 shows few nominalizations, passives, or prepositional phrases; and it contains numerous references to the temporal/physical situation (*right over on the stand side, in the centre, to the front, way back*). Thus consideration of both Dimensions 1 and 2 is required to show that broadcasts are like many written text types in permitting little interaction or personal involvement—but are like many spoken text types in having a concrete, highly situated content.

Finally, a description of the characteristics of fiction would be inadequate if only one dimension were considered. Along Dimension 1, fiction occurs toward the lower end with most of the written text types, but has a higher score than the expository types (excluding professional letters); it is thus highly explicit in its expression of meaning (having no real-time production con-

straints), but highly interactive, with considerable personal involvement (cf. Tannen 1982a). The high interaction may be related, in part, to the inclusion of fictional dialog in the text samples of the present study. Along Dimension 2, fiction gets very low scores, grouping with the majority of the spoken types. This shows the lack of abstract content in fiction (note the absence of passives and nominalizations in Sample 3),[12] as well as its highly situated nature (note the many references to a physical/temporal situation in Sample 3: *now, into the dugout, coming in from left field, beside it, right up to the dugout*). Finally, along Dimension 3, fiction is shown to be different from all other spoken/written types by its strong preference for reported events over either description of events actually in progress or expository presentation of information.

Other such comparisons could be made. In all these cases, text types that are similar along one dimension are shown to be quite distinct with respect to other dimensions. An over-all model of the relations among written/spoken types requires simultaneous consideration of the operation of all three textual dimensions.

Reconciliation of conflicting claims

5. No single explanation can be given for the confused picture of the relationship between speech and writing which has emerged from past research. In §1, three restrictions were identified which characterize many earlier quantitative studies: paucity of linguistic measures, of text samples, and of text types. In addition, the statistical methods which were typically used—direct comparison of frequencies, mean scores, or multiple univariate tests of significance—restricted the results. These restrictions were essential for exploratory research, where detailed analysis of individual linguistic features in individual texts was required. The results obtained in the present study would not have been interpretable without the detailed findings of earlier studies. Furthermore, the wide scope of the present study, including 545 texts taken from 16 text types, is possible only because of the availability of large-scale computerized corpora and computer-aided analysis. As it turns out, however, restricted, detailed studies do not provide an adequate basis for global conclusions concerning the relationship between speech and writing. In fact, the combination of the above-mentioned restrictions essentially guaranteed contradictory global conclusions. The restriction to few linguistic features and text types precludes the discovery of multiple textual dimensions, and the use of univariate statistics will fail to discover variation along multiple dimensions, even if the text sample includes such variation. Thus contradictory global con-

[12] The absence of features like passives and nominalizations in fiction (as well as their presence in some of the other text types) is probably related to social and attitudinal parameters in addition to the abstractness of the content. Thus the use of passives and nominalizations could be seen as establishing a learnèd, highly impersonal style, which can be used to maintain social distance between author and reader (or speaker and listener). Students are trained to avoid these features in writing, although the effect of this training is not clear. The absence of these features in fiction, though, is probably related to these parameters as well as to content type.

clusions have frequently resulted from comparison of different text types with respect to linguistic measures taken from different textual dimensions.

The resolution of contradictory global conclusions can be illustrated through a comparison of Chafe 1982 and Blankenship 1962. These admittedly preliminary studies (see Chafe & Danielewicz 1985, Blankenship 1974) are used here because their findings are illustrative of the problem. Chafe found large linguistic differences between speech and writing by comparing conversation and academic papers along the two hypothesized dimensions of integration/fragmentation and involvement/detachment. By contrast, Blankenship 1962 concluded, from a comparison of popular journal articles and public speeches, that there was no over-all difference between speech and writing with respect to the measures of sentence length, past tense forms, and passives.

Chafe's text types strongly biased his study in favor of finding a spoken/written distinction, since conversation and academic prose show polar distinctions along Dimensions 1 and 2 identified in the present study: Interactive vs. Edited Text, and Abstract vs. Situated Content. If conversation is taken to represent speech, and academic prose to represent writing, then most linguistic features considered in previous research could be presented as evidence for a spoken/written distinction.

Blankenship compared text types which were more similar (popular journals and prepared speeches)—but with respect to linguistic features which, as I have shown here, are taken from separate textual dimensions: past tense clusters with perfect aspect and 3rd person pronouns on Dimension 3 (Reported vs. Immediate Style), while passives cluster with nominalizations and prepositional phrases on Dimension 2. Blankenship found nearly the same number of passives in writing and speech, and more past-tense forms in speech. These results are confirmed by the findings in §4, above: popular lore and planned speeches have nearly the same value on Dimension 2, and planned speeches show a higher score than popular lore on Dimension 3. However, we cannot accept Blankenship's global conclusion that the reverse results for these two features demonstrate the unimportance of the spoken/written distinction. That is, the relations among text types differ along each of the textual dimensions described above in §4; and the only major distinction between the two text types used by Blankenship is along Dimension 1—which is not represented in her selection of linguistic features.

More specific contradictions can also be resolved with the over-all model developed here. Thus Chafe found more passives in writing (academic papers) than in speech (conversation); Poole & Field, however, found few passives in either writing (narratives) or speech (interviews). The distribution of these text types along Dimension 2, which includes passive constructions, confirms both pairs of findings: academic prose has one of the highest scores, and conversation one of the lowest; fiction (the text type in the present study most similar to 'narrative') has an intermediate score, as do interviews.

Contradictory findings concerning the extent of subordination in speech and writing are more difficult to resolve because the function of subordination is more varied and complex than these other linguistic features. In terms of the

over-all amount of subordination in a text, Kroll, O'Donnell, and Chafe found more in writing; but Halliday and Poole & Field found more in speech. Beaman, who generally found more subordination in speech, perceptively noted that different types of subordination are present in the two modes.[13] The differences among these findings are influenced by several parameters. For example, the text types chosen for comparison vary widely between studies: while O'Donnell looked at only one television interview and one newspaper opinion column, Chafe compared conversation and academic papers, and Beaman compared spoken/written narratives.

Equally important is the fact that the subordination measures used in earlier studies are different and not directly comparable. As I have shown, *that*-clauses, *if*-clauses, WH-clauses, and other subordinators (i.e. adverbial clauses) function as part of a single dimension; but relative clauses have a separate communicative function. Infinitives have been grouped as parts of two different factors, and thus may have yet another communicative function. Assessments of over-all subordination which indiscriminately lump these measures together can be expected to produce contradictory results.

When individual subordination measures are considered separately, the findings are less contradictory; e.g., in agreement with my analysis, most previous studies have found more *that*-clauses in speech (Beaman 1984, Frawley 1982, Jörgensen 1978, and even O'Donnell 1974). However, nearly all studies have shown relative clauses to occur more frequently in writing (Frawley, Kroch & Hindle, O'Donnell, Chafe); they are distributed differently from *that*-clauses, as shown by their clustering on separate dimensions in the present analysis. Finally, although O'Donnell and Beaman found more adverbials in writing (an apparent contradiction to the clustering of 'other adverbial subordinators' on Factor 1 in my analysis), Beaman's specific findings for speech agree well with mine: more condition adverbials (comparable to *if*-clauses here) and reason adverbials (e.g. *though, although, because*, and *since*—which are several of the primary tokens in my category 'other subordinators').

We have seen that subordination features function as part of different textual dimensions, and that they serve differing functions in different text types; contrast *that*-clauses on Dimension 1 with relative clauses on Dimension 4. More detailed study of subordination features as they function in different text types is required before final conclusions can be drawn concerning their over-all distribution and functions.

This section has shown that the global conclusions reached in previous research are contradictory because the text types chosen for comparison were too similar or too different; because the linguistic features chosen belonged to different textual dimensions; and because researchers relied on inadequate analytical techniques. The analysis presented in §4, above, more adequately represents the complex relations among spoken/written text types in English, and

[13] Studies in other languages indicate similar distributions. Pellegrino & Scopesi 1978 found the same over-all amount of subordination in spoken and written Italian, as did Jörgensen 1978 for Swedish.

shows that the 'contradictions' are in fact accurate findings. That is, individual findings about text types have been generally valid, but the global conclusions based on them, regarding speech and writing, have been in error.

Summary and conclusion

6. This study has identified three fundamental parameters of textual variation which underlie speech and writing in English. These dimensions are tentatively labeled 'Interactive vs. Edited Text', 'Abstract vs. Situated Content', and 'Reported vs. Immediate Style'. In addition, the similarities and differences among 16 spoken/written text types with respect to these dimensions have been charted. This treatment depends on a multi-feature/multi-dimension approach, using factor analysis to study a wide range of linguistic features taken from diverse text types. Earlier exploratory studies, using a more restricted analysis of few text types and few linguistic features, reached contradictory global conclusions concerning the relations among spoken/written text types. Each of these studies accurately described a small piece of the over-all puzzle; however, comprehensive conclusions regarding the over-all relations among spoken/written text types require a comprehensive methodology. It is shown in §4, above, that inaccurate descriptions of the relations among text types result if consideration is limited to only one or two of the dimensions identified here. It is shown further, in §5, that the contradictory global conclusions reached in previous studies can be attributed to imbalance in the text types being compared, and to the use of linguistic features belonging to different textual dimensions. The value of the present approach is shown by the identification of three textual dimensions, which can be used to verify the individual findings of earlier studies while resolving the contradictory global conclusions reached in those studies.

A multi-feature/multi-dimension approach has proved equally valuable to other areas of text analysis. It has been used to identify dimensions of linguistic complexity (Finegan & Biber 1984), sociolinguistic prestige (Finegan & Biber 1983), style (Grabe 1984, Biber & Finegan 1984), dialect differences (Biber 1984), and stance (Biber & Finegan 1985); it has also been used to identify dimensions of formality and literacy in a non-Western language (Besnier 1984). There are doubtless other applications as well. Textual variation across any sociolinguistic domain is complex and multi-faceted; hence comprehensive conclusions concerning variation within such domains must be based on a comprehensive methodology. The present study uses such methodology to identify three principal textual dimensions underlying variation among spoken/written text types in English.

Appendix

List of the 41 syntactic and lexical features used, with interpretations suggested in previous studies.

Writing has a more detached, decontextualized style:

1. Nominalizations (Chafe 1982, DeVito 1967).
2. Agentless passives (Weiner & Labov 1983).
3. *By*-passives (Blankenship 1962, Brown & Yule 1983, Chafe 1982, Chafe & Danielewicz 1985, O'Donnell 1974).

Writing has a more elaborated, expanded style:

4. *That*-clauses (Beaman 1984, O'Donnell 1974, Chafe 1982, Poole & Field 1976, Frawley 1982).
5. Relative clauses (Beaman, O'Donnell, Chafe, Poole & Field, Frawley).
6. Infinitives (Beaman, Chafe, O'Donnell).
7. Other adverbial subordinators: *(al)though, because, since, unless, while, whilst, until, in that, so that, such that* (Beaman, O'Donnell).
8. WH-clauses.
9. *If*-clauses.
10. Total prepositional phrases (Blankenship 1974, Chafe 1982, Chafe & Danielewicz 1985, Marckworth & Baker 1974).
11. Total adjectives (Blankenship 1974, Chafe 1982, O'Donnell 1974, Marckworth & Baker, Poole & Field).
12. Other adverbs—excluding place and time adverbs, hedges, emphatics, style and attitudinal disjuncts (Blankenship 1974, Marckworth & Baker, Poole & Field, Tottie 1983).

Writing has a more explicit level of expression:

13. Type/token ratio in vocabulary (Blankenship 1974, Drieman 1962, Chafe & Danielewicz).
14. Word length (Blankenship 1974, Drieman 1962).
15. Specific conjuncts: *in conclusion, in consequence, hence* etc. (Tottie 1983, Quirk et al. 1972).
16. Style disjuncts: *bluntly, flatly, truly* etc. (Quirk et al., Biber & Finegan 1985).
17. Attitudinal disjuncts: *admittedly, assuredly, certainly* etc. (Quirk et al., Biber & Finegan 1985).
18. Specific hedges (and down-toners): *virtually, possibly, perhaps* etc. (Chafe & Danielewicz, Quirk et al.).
19. Specific emphatics (and intensifiers): *completely, absolutely, precisely* etc. (Quirk et al.).

Writing has more explicit marking of informational relations:

20. *It*-clefts (Brown & Yule, Marckworth & Baker).
21. WH-clefts (Pawley & Syder 1985, Marckworth & Baker).
22. Demonstrative-BE-WH (Quirk et al.).

Speech has a more inexplicit, informal, style of expression:

23. General (informal) hedges: *at about, something like, sort of* etc. (Chafe, Chafe & Danielewicz).
24. General (informal) emphatics: *just, really, real* + ADJ, *so* + ADJ etc. (Chafe, Brown & Yule).
25. Pronoun *it* (Chafe & Danielewicz).
26. Pro-verb *do*.
27. Contractions (Chafe & Danielewicz, Marckworth & Baker).
28. Subordinator-*that* deletion: *I think* [*that*] *he went to* ... (Beaman 1984, Finegan & Biber 1984).
29. Relative pronoun deletion: *the speech* [*that*] *I heard was* ... (Beaman 1984, Finegan & Biber 1984).
30. Final (stranded) prepositions (Finegan & Biber 1983).
31. Split infinitives (Finegan & Biber 1983).
32. Split auxiliaries: AUX ADV VERB (Finegan & Biber 1983).

Speech is more interactive/involved:

33. First and second person pronouns (Chafe 1982, Chafe & Danielewicz, Blankenship 1974, Poole & Field).
34. Direct yes–no questions (Marckworth & Baker).
35. Direct WH-questions (Marckworth & Baker).

Speech is more situated in a physical/temporal context:

36. Place adverbs: *across, below, outside* etc. (Chafe & Danielewicz).
37. Time adverbs: *early, late, soon* (Chafe & Danielewicz).
38. Third person pronouns (Poole & Field).

Speech and writing differ in their use of verb tense/aspect:

39. Past tense (Blankenship 1962, Marckworth & Baker).
40. Perfect aspect (Marckworth & Baker).
41. Present tense.

REFERENCES

AKINNASO, F. NIYI. 1982. On the differences between spoken and written language. Language and Speech 25.97–125.

BEAMAN, KAREN. 1984. Coördination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. Coherence in spoken and written discourse, ed. by Deborah Tannen, 45–80. Norwood, NJ: Ablex.

BESNIER, NIKO. 1984. Restricted literacy and the development of a literacy–orality contrast. Dissertation proposal, Department of Linguistics, University of Southern California.

BIBER, DOUGLAS. 1984. A textual comparison of British and American writing. Paper presented at NWAVE 13, Philadelphia.

——. 1985. Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. Linguistics 23.337–60.

——, and EDWARD FINEGAN. 1984. Luddites, linguists or the literati: Who speaks to style in the computer age. Paper presented at Life of the Mind Series, Division of Humanities, University of Southern California.

——, ——. 1985. Adverbials as markers of stance: A multivariate approach. Paper presented at 6th Intl. Conf. on Eng. Lang. Research on Computerized Corpora, Lund, Sweden.

BLANKENSHIP, JANE. 1962. A linguistic analysis of oral and written style. Quarterly Journal of Speech 48.419–22.

——. 1974. The influence of mode, submode, and speaker predilection on style. Speech Monographs 41.85–118.

BROWN, GILLIAN, and GEORGE YULE. 1983. Discourse analysis. Cambridge: University Press.

CARROLL, JOHN B. 1960. Vectors of prose style. Style in language, ed. by Thomas A. Sebeok, 283–92. Cambridge, MA: MIT Press.

CAYER, ROGER L., and RENEE K. SACKS. 1979. Oral and written discourse of basic writers: Similarities and differences. Research in the Teaching of English 13.121–8.

CHAFE, WALLACE L. 1982. Integration and involvement in speaking, writing, and oral literature. In Tannen 1982c, 35–54.

——, and JANE DANIELEWICZ. 1985. Properties of spoken and written language. Comprehending oral and written language, ed. by Rosalind Horowitz and S. J. Samuels, to appear. New York: Academic Press.

DEVITO, JOSEPH A. 1967. Levels of abstraction in spoken and written language. Journal of Communication 17.354–61.

DRIEMAN, G. H. J. 1962. Differences between written and spoken language. Acta Psychologica 20.36–57, 78–100.

DUBUISSON, COLETTE; LOUISETTE EMIRKANIAN; and DAVID SANKOFF. 1983. The development of syntactic complexity in narrative, informative and argumentative discourse. Montréal: Centre de Recherche de Mathématiques Appliquées, Université de Montréal.

FAIGLEY, LESTER, and PAUL MEYER. 1983. Rhetorical theory and readers' classifications of text types. Text 3.305–25.

FINEGAN, EDWARD. 1985. The linguistic forms of prestige. The miracle of language: Essays in honor of Charlton Laird, ed. by Robert Gorrell & Phillip Boardman. To appear, Reno: University of Nevada Press.

——, and DOUGLAS BIBER. 1983. Towards a unified model of sociolinguistic prestige. Proceedings of NWAVE 12, ed. by David Sankoff, to appear.

——, ——. 1984. A multi-dimensional model of linguistic complexity in English. Social and cognitive perspectives on language, ed. by Christopher J. Hall et al. (Southern California papers in linguistics, 11.) To appear, Los Angeles: University of Southern California.

FRANCIS, W. NELSON, and HENRY KUČERA. 1982. Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton Mifflin.

FRAWLEY, WILLIAM. 1982. Universal grammar and composition: Relativization, complementation, and quantification. Linguistics and literacy, ed. by W. Frawley, 65–90. New York: Plenum.
FREDERIKSEN, CARL H., and JOSEPH F. DOMINIC. 1981 (eds.) Writing: The nature, development, and teaching of written communication, II: Process, development and communication. Hillsdale, NJ: Erlbaum.
GORSUCH, RICHARD L. 1983. Factor analysis. 2nd edn. Hillsdale, NJ: Erlbaum.
GRABE, WILLIAM P. 1984. Towards defining expository prose within a theory of text construction. USC dissertation.
GREEN, GEORGIA M., and JERRY L. MORGAN. 1981. Writing ability as a function of the appreciation of differences between oral and written communication. In Frederiksen & Dominic, 177–88.
GUMPERZ, JOHN J.; HANNAH KALTMAN; and MARY CATHERINE O'CONNOR. 1984. Cohesion in spoken and written discourse. Coherence in spoken and written discourse, ed. by Deborah Tannen, 3–20. Norwood, NJ: Ablex.
HALLIDAY, MICHAEL A. K. 1979. Differences between spoken and written language: Some implications for literacy teaching. Communication through reading: Proceedings of the 4th Australian Reading Conference, ed. by Glenda Page et al., 2.37–52. Adelaide: Australian Reading Association.
HORVATH, BARBARA M. 1985. Variation in Australian English: The sociolects of Sydney. Cambridge: University Press.
HYMES, DELL. 1972. On communicative competence. Sociolinguistics, ed. by John B. Pride & Janet Holmes, 269–93. Harmondsworth: Penguin.
JÖRGENSEN, NILS. 1978. Underordnade satser och fraser i talad svenska: Funktion och byggnad. (Lundastudier i nordisk sprakvetenskap, C-10.) Lund: Ekstrand.
JOHANSSON, STIG. 1982 (ed.) Computer corpora in English language research. Bergen: Norwegian Computing Centre for the Humanities.
——; GEOFFREY N. LEECH; and HELEN GOODLUCK. 1978. Manual of information to accompany the Lancaster–Oslo/Bergen Corpus of British English, for use with digital computers. Oslo: Department of English, University of Oslo.
KAY, PAUL. 1977. Language evolution and speech style. Sociocultural dimensions of language change, ed. by Ben G. Blount & Mary Sanches, 21–33. New York: Academic Press.
KROCH, ANTHONY S., and DONALD M. HINDLE. 1982. A quantitative study of the syntax of speech and writing. Final report to the National Institute of Education.
KROLL, BARBARA. 1977. Ways communicators encode propositions in spoken and written English: A look at subordination and coördination. Discourse across time and space, ed. by Elinor O. Keenan & Tina Bennett (Southern California occasional papers in linguistics, 5), 69–108. Los Angeles: USC.
MARCKWORTH, MARY L., and WILLIAM J. BAKER. 1974. A discriminant function analysis of co-variation of a number of syntactic devices in five prose genres. American Journal of Computational Linguistics, Microfiche 11.
O'DONNELL, ROY C. 1974. Syntactic differences between speech and writing. American Speech 49.102–10.
OLSON, DAVID R. 1977. From utterance to text: The bias of language in speech and writing. Harvard Educational Review 47.257–81.
——, and NANCY TORRANCE. 1981. Learning to meet the requirements of written text: Language development in the school years. In Frederiksen & Dominic, 235–55.
PAWLEY, ANDREW, and FRANCES H. SYDER. 1985. Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. Journal of Pragmatics, to appear.
PELLEGRINO, M. L. MORRA, and A. A. SCOPESI. 1978. Oral and written language in children: Syntactical development of descriptive language. International Journal of Psycholinguistics 5.5–19.
POOLE, MILLICENT E. 1973. A comparison of the factorial structure of written coding

patterns for a middle-class and a working-class group. Language and Speech 16.93–109.
——. 1979. Social class, sex and linguistic coding. Language and Speech 22.49–67.
——, and T. W. Field. 1976. A comparison of oral and written code elaboration. Language and Speech 19.305–11.
Quirk, Randolph; Sidney Greenbaum; Geoffrey Leech; and Jan Svartvik. 1972. A grammar of contemporary English. London: Longman.
Rader, Margaret. 1982. Context in written language: The case of imaginative fiction. In Tannen 1982c, 185–98.
Redeker, Gisela. 1984. On differences between spoken and written language. Discourse Processes 7.43–55.
Rubin, Andee. 1980. A theoretical taxonomy of the differences between oral and written language. Theoretical issues in reading comprehension, ed. by Rand J. Spiro et al., 411–38. Hillsdale, NJ: Erlbaum.
Svartvik, Jan, and Randolph Quirk. 1980 (eds.) A corpus of English conversation. Lund: Gleerup.
Tannen, Deborah. 1982a. Oral and literate strategies in spoken and written narratives. Lg. 58.1–21.
——. 1982b. The myth of orality and literacy. Linguistics and literacy, ed. by William Frawley, 37–50. New York: Plenum.
——. 1982c (ed.) Spoken and written language: Exploring orality and literacy. Norwood, NJ: Ablex.
Tottie, Gunnel. 1983. The missing link? Or, why is there twice as much negation in spoken English as in written English? Stockholm Studies in English 62 (Proceedings from the 2nd Scandinavian Symposium on Syntactic Variation, ed. by Sven Jacobson), 67–74.
——. 1984. Is there an adverbial in this text? (And if so, what is it doing there?) Proceedings from the 2nd Nordic Conference for English Studies, ed. by Hakan Ringbom & Matti Rissanen, 299–315. Åbo, Finland: Åbo Akademi Foundation.
——; Bengt Altenberg; and Lars Hermeren. 1983. English in speech and writing. Lund: Engelska Institutionen.
Weiner, E. Judith, and William Labov. 1983. Constraints on the agentless passive. Journal of Linguistics 19.29–58.
Zipf, George K. 1949. Human behavior and the principle of least effort. Cambridge, MA: Addison-Wesley.

[Received 20 February 1985;
revision received 14 June 1985;
accepted 8 July 1985.]