Lenguas Modernas 18 (1991), 133-160
Universidad de Chile.

MULTIVARIATE APPROACHES TO CONTRASTIVE RHETORIC

PAUL LUX AND WILLIAM GRABE
Texas Tech University and Northern Arizona University

The purpose of this article is to describe applications of multivariate quantitative research approaches to contrastive rhetoric research. We will first introduce the notion of contrastive rhetoric and evidence supporting the notion. We will then present the rationale for the multivariate approach used here. Two separate studies using a multivariate approach will then be described. One examines patterns of variation in English and Brazilian Portuguese newspaper editorials (Dantas-Whitney and Grabe to appear); the second examines variations in writing among Ecuadorian Spanish and Anglo-American English university students (Lux 1991). Results of these studies suggest that a multivariate approach to text analysis, and specifically to contrastive rhetoric, is a productive line of research.

1. CONTRASTIVE RHETORIC

The notion that different cultural and linguistic groups will exhibit different preferences for the rhetorical organization of written texts has been open to debate since Kaplan's (1966) now-famous article in *Language Learning*. While the evidence supporting such an argument is not easy to demonstrate, and much earlier work on contrastive rhetoric has been criticized on this ground, the notion of contrastive rhetoric has a strong intuitive appeal, as most second language teachers of writing can readily attest (Kaplan 1988, Swales 1990). Over the last years, however, a wide array of research, both outside and within written text analysis, can be assembled which make a strong set of arguments for the notion of contrastive rhetoric.

1.1 *External research evidence*: Supporting evidence for contrastive rhetoric comes first from psycholinguistics with the cross-linguistic language development research of Slobin and his colleagues (e.g., Berman 1985, Slobin 1987, Slobin and Bocaz 1988). In this line of research, Slobin and others have noted how children acquiring different languages will exhibit preferences for different sorts of linguistic structures, structures which are prevalent in each of the languages examined. The structural differences uncovered do not argue for a strong form of the Whorfian hypothesis, but for the recognition of cultural preferences which make greater use of certain options among linguistic possibilities. (See also Hunt and Agnoli 1991.)

Sociolinguistics and the study of literacy have provided further evidence of variation in discourse which can only be understood in terms of its sociocultural contexts; that is, patterns of discourse use are socially and culturally shaped. The literacy research of Heath (1983, 1986a, 1986b), Shor (1987), and Street (1984) provide extensive evidence to support this position. Additional evidence for the socially grounded preferential uses of discourse routines comes from research in classroom contexts, as can be seen in a variety of studies (Boggs 1985, Cazden 1988, Collins and Michaels 1986, Scollon and Scollon 1981). Outside of the specific context of the classroom, the sociolinguistic exploration of language socialization, and the influence of such socialization on patterns of language use support a contrastive rhetoric perspective (e.g., Besnier 1988, Boggs 1985, Clancy 1986, Ochs 1988, Phillips 1983, Schieffelin and Gilmore 1986, Schieffelin and Ochs 1986). Different cultures have different ways of doing things with language. These different uses are culturally/socially shaped and they have reflexes in the organization of discourse. It would seem reasonable to suggest that the same social context influences on the organization of oral-language use would arise as well in written discourse.

The development of social construction theory and its influence on rhetoric as discussed by Bazerman (1988, 1991), Bruffee (1986), Kaplan and Grabe (1991), and Myers (1985, 1990) provide yet another source of evidence for a contrastive rhetoric perspective. In particular, the study of scientific discourse in the last decade has demonstrated that even the most "objective" and universal of discourse genres is, in fact, a product of rhetorical socialization and preferred conventionalizations accepted by the scientific community (Atkinson 1991, Swales 1990). Scientists do not write objective value-neutral articles; they respond to the current trends in theory making and shape their reporting accordingly (much as this article is doing).

A final area supporting the notion of contrastive rhetoric can be seen in rhetoric with the emergence of post-structural approaches to the critical study of text and their emphases on the socio-historical forces which shape our writing and our reading of any text (Crowley 1990, Derrida 1976, Neel 1988). The notion that any text could be objective, free from contextual shaping, does not appear logically valid. (See also the research of Critical Discourse Analysis: Birch 1989, Fairclough 1989, Kress 1989, 1991, Wodak 1989.) In the context of contrastive rhetoric, variation attributable to distinct cultures and languages would inevitably cause different contextual shaping and different preferential patterns of textual organization.

Given this diverse array of evidence for the general notion of contrastive rhetoric, it would seem appropriate to acknowledge its broad theoretical appeal. At the same time, it is important to consider the range of evidence and supporting arguments emerging specifically from the direct study of contrastive rhetoric across many languages. As the next section indicates, there is now considerable written discourse research supporting the notion of contrastive rhetoric (Grabe and Kaplan 1988, Leki 1991).

1.2. *Text-based research evidence:* The recent evolution of research in contrastive rhetoric is first indicated by the increasing stress being placed on the comparison of texts in two different languages rather than primarily on the English writing of ESL/EFL

students. (See Purves and Hawisher 1990 for an interesting comparison across many languages.) In many recent studies, the focus is on direct comparison of syntactic and discourse features in texts of two or more languages. Many specific language comparisons (primarily with English) have been made following this line of research.

Clyne (1983, 1985, 1987) has argued that different organizational preferences exist in English and German 'learned' prose. Hinds (1983, 1987, 1990) has argued that Japanese has a major form of textual organization with no direct English counterpart *(Ki-Shoo-Ten-Ketsu)*. Similar patterns have been argued for in Korean (Eggington 1987) and Chinese (Cheng 1985). Hinds (1987, 1990) has also argued that reader-writer relationships in Japanese prose are different from those in English prose, and that a number of Asian languages have a distinct 'quasi-inductive' prose style. Bickner and Peyasantiwong (1988) and Indrasuta (1988) have argued for differences in Thai and English essays based on comparisons of Thai students writing in Thai and English, as well as U.S. students writing in English. Kachru (1983, 1988) has argued that Hindi expository prose is distinct from English in that a spiral prose organization is discernable (based on traditional Hindi text organization). She also argues that traditional Hindi organizational patterns are found in Indian English writing. Both Sa'Adeddin (1988) and Zellermayer (1988) argue that Semitic languages (Arabic and Hebrew, respectively) demonstrate preferences for oral language features of texts such as additive conjunctions, repetition, and implicit references. (See also Ostler 1987.) Montano-Harmon (1988, 1991) has argued that additive conjunction relations are used more extensively in Spanish than they are in English in her comparison of Mexican writers in Spanish and Anglo writers in English.

In addition to the wide range of studies involving the comparison of two languages, a growing body of discourse based research has led to the study of syntactic/textual features central to the structuring of discourse (Biber 1988, Carlson 1988, Connor 1990, Connor and Johns 1990, Grabe 1987, 1990, Reid 1988.) A number of these findings have been incorporated into recent research on contrastive rhetoric. In addition, recent research in discourse analysis has extended the study of text to include analyses which assume larger divisions in the organization of text. Earlier efforts have included Kaplan's (1972, Ostler 1987) Discourse Bloc framework. More recent approaches include measures of persuasive discourse (Connor 1990, Connor and Lauer 1988), storygraph analysis (Soter 1988), topical structure analysis (Cerniglia et al. 1990, Lauttamatti 1987), and markers of superstructure (Evenson 1990). Finally, discourse research has begun to focus on textual analyses which are not directly dependent on individual linguistic features of the text (Besnier 1988, Biber 1988, Grabe 1987). The rest of this paper will explore this final research approach as it applies to comparisons between Brazilian Portuguese and English, and Ecuadorian Spanish and English.

2. A MULTIDIMENSIONAL APPROACH TO TEXT ANALYSIS

The studies reported here both draw upon a corpus-based approach to text analysis which was first developed by Biber and termed as a Multi-Feature Multi-Dimensional Analysis (1985, 1988, 1991). Biber's research on oral and written discourse has been pioneering to the field of discourse analysis, overcoming many problems in previous discourse studies which examine a text-based corpus. His research argues that much

of the confusing results from earlier studies of spoken-written variation in discourse centered on at least one of the following methodological concerns: the inappropriate comparison of different types of texts, the examination of only a limited number of texts, and the assumption that specific surface features could, a priori, be fully representative of communicative dimensions of texts.

Perhaps the most important result of Biber's research on spoken-written variation is the demonstration that text analysis cannot assume a single theoretical dimension along which all surface features of a text will vary. Rather, Biber has shown that any text is a reticulated structure, being composed of a number of communicative dimensions. Each dimension will vary from text to text and from text type to text type — simply put, texts are multidimensional structures and must be analyzed as such. The notion that texts are multidimensional constructs is not a claim attributable solely to Biber, and it is now a well-accepted tenet of discourse analysis (e.g., Chafe 1982, Kaplan 1987, Tannen 1982).

A less well-acknowledged assumption of Biber's approach is that surface linguistic features will contribute to define communicative textual functions, in particular by the co-occurrence patterns of many of these linguistic features. This assumption, however, is supported by current sociolinguistic theory. As Biber (1988:21) notes:

> Ervin-Tripp (1972) and Hymes (1974) discuss co-occurrence relations among linguistic features in terms of 'speech styles', a variety or register that is characterized by a set of co-occurring linguistic features... Although the theoretical importance of co-occurrence patterns among linguistic features has been well-established by these researchers, the empirical identification of salient co-occurrence patterns in English discourse has proven to be difficult.

A second research approach which supports this perspective derives from British work on text analysis. Hoey (1991), in a study of lexis relations in written discourse, demonstrates that types of lexical co-occurrence patterns contribute to textual coherence of non-narrative prose over and above any specific set of lexical forms. Moreover, the general functional linguistic approach developed by Halliday (1985, Halliday and Hasan 1989) argues for complex co-occurrence patterns as indications of basic form-function relations in discourse.

2.1. *Biber's 1988 study:* Biber's (1988) research sought to avoid the shortcomings associated with prior corpora research. In his study of spoken and written variation, he examined 67 different linguistic features in 481 different texts of 23 different types, including academic prose, general fiction, biographies, press reportage, popular lore, professional letters, face-to-face conversation, broadcasts, and planned speeches. The texts, comprising approximately one million words in all, were analyzed by computer to identify and count the relevant linguistic features. The quantitative results were then used in a factor analysis (Principal Factor Analysis, promax rotation). The use of factor analysis in this research allows for *the identification of co-occurring linguistic features which may indicate a shared communicative function;* that is to say, the factor dimensions can be interpreted in terms of the communicative functions most widely shared by the co-occurring linguistic features (Biber 1986, 1988). Overall, the 67 features Biber (1988) examined fell together into seven factors (See Appendix 1.) Six of these factors proved to be highly reliable and are listed below:

1. Involved versus Information Production
2. Narrative versus Non-narrative Concerns
3. Explicit versus Situation-Dependent Reference
4. Overt Expression of Persuasion versus Non-overt Expression
5. Abstract versus Non-abstract Information
6. On-line Informational Elaboration

A simplified explanation of Biber's first dimension will illustrate how he analyzed the factor dimensions. Biber's first dimension is interpreted as "Involved versus Informational Production." He describes the positively-valued linguistic features which co-occur on this dimension as "marking a high level of interaction and personal affect" as well as "a generalized and fragmented presentation of content;" features with negative values are described as being "highly informational [and showing] almost no concern for interpersonal involvement or affective content." Strong positive-oriented linguistic features included *first* and *second person pronouns, private verbs, emphatics,* and *WH questions* for involvement; *hedges, discourse particles, contractions, non-phrasal and* and the *pronoun it* for fragmented content. Negative linguistic features included *high frequency of nouns, high frequency of prepositions, few verbs, long words,* and a *high type/token ratio* for non-involvement. Discussions in the research literature provided the motivation for interpreting the co-occurrence of these features in terms of the dimension labels given. This approach was repeated for interpreting each of the other dimensions from the factor analysis. The results of Biber's (1988) study demonstrated that it is unrealistic to compare (spoken and written) texts based only on one dimension, and that at least six dimensions *may* be looked at in comparative text analysis, including research in contrastive rhetoric.

2.2. *Grabe's 1987 study:* In a study with a similar general methodology, Grabe (1987) examined text type variation in written prose. The primary purpose was to examine whether or not it is possible to define the term *expository prose* as a text type based on linguistic feature co-occurrence and textual dimensions. The study demonstrated that expository prose is a valid genre label and that a number of sub-genres may be defined as well. In addition, this study confirmed a number of Biber's findings using an independent corpus of texts, and involving a different range of text variation (e.g., no spoken texts used in the corpus). More importantly for the present studies, it indicated a way to define text comparability in two or more languages by developing a model of textual dimensions in written prose.

Concentrating only on expository prose, Grabe analyzed 150 texts, of 15 functional types, in terms of 31 syntactic, lexical, and cohesion variables. Using Biber's program for counting the linguistic features, Grabe also used factor analysis to derive a seven-factor solution. The first six factors proved to be highly reliable and were given interpretations. These interpretations were based on the important co-occurrence patterns of the linguistic features on each factor dimension.

Figure 1 provides the factor structure for the first four factors, since they were the strongest and most amenable to measurement, as well as the important linguistic variables which "loaded" on each factor. Factor one, for example, was created by the

co-occurrence of the following features in the corpus: *Present Tense, Precise Conjuncts, Relative Clauses,* and *Comparative Cohesion Relations;* and negatively by *Past Tense.*

Factor 1	Factor 2	Factor 3	Factor 4
Immediacy of Context	Orientation of Discourse	Information Type	Presentation of Information
Immediate Present tense (.7538) Precise conj (.4963) Relative cl (.3858) Comparatives (.3871)	Interactional General hedge (.3747) I/You (.7411) Infinitives (.5210) That clauses (.3554) Subordination (.4012) Pro-verb do (.4576) Present tense (.4141)	Abstract Nominatives (.5311) Prepositions (.6358) Words/Sentence (.6052)	Objective Passives (.5373) Repetition (.6348)
Distant Past tense (-.8212)	Informational Repetition (-.3500)	Situational He/She (-.3848) Locative adv (-.3473)	Expressive He/She (-.6595) Synonymy/ant (-.3654)

Figure 1: Factor structure interpretations from Grabe (1987)
(First four factors)

That is, the positive and negative features combined in similar ways in many texts to create dimension one. This dimension was interpreted as an 'Immediacy of Context' dimension. One reasonable interpretation of this dimension in terms of general genre distinctions is to see 'Immediacy of Context' as a basic division of narrative texts (information that was) and descriptive texts (information that is). While the use of description is likely to run through many texts, it is plausible to argue that most descriptions use present tense, *Be* as main verb, and relative clauses more than other genres (Martin 1989, Swales 1990). Thus, it is possible to interpret dimension one as opposing the use of present time description with past time narration. To the extent that most expository prose text types use description more than narrative, this dimension could also be viewed as a Non-narrative–Narrative dimension.

Following Biber's (1988) situational and communicative interpretations for factor dimensions, the four important dimensions in Figure 1 were labelled as follows:

1. Non-Narrative versus Narrative Context
2. Interactional versus Informational Orientation

3. Abstract/Logical versus Situational Information
4. Objective versus Expressive Style

These four dimensions served to define variation among the 15 text types used in the Grabe study. The importance of this research for contrastive rhetoric is that it suggests a way to examine prose text types in more than one language while allowing for a multidimensional text analysis. In order to test out the feasibility of this notion, we applied this methodology to two contrastive analyses: the first is a comparison of Portuguese and English editorial texts; the second is a comparison of essays written by Ecuadorian and U.S. university students.

3. Comparing Brazilian Portuguese and English editorial texts

3.1. *Background:* This section is based on an exploratory research study carried out by Dantas-Whitney and Grabe (to appear) comparing editorial texts in Brazilian Portuguese and English. Twenty texts were chosen for analysis and were compared for 15 linguistic variables taken from the first three factor dimensions in Grabe's (1987) study; these 15 linguistic features provided a way to explore the applicability of a multivariate approach to contrastive analysis. Since no oral language data was involved, it was felt that the dimensions created by the range of expository prose texts in Grabe's study would be more relevant than the factor dimensions in Biber (1988). In addition to exploring the application of Biber's approach, the purpose of this study was to see whether or not the Brazilian Portuguese editorials would vary in significantly different ways from the English editorial texts along the three textual dimensions already established for English expository prose.

3.2. *Method:* The texts used in this study consisted of ten texts from the editorial page of the *Christian Science Monitor* and ten texts from the editorial page of *Journal do Brasil.* These two newspapers were considered equivalent in relation to their circulation characteristics and reputation in the US and in Brazil. The editorials were matched according to the number of words, subject matter, and approximate date of publication. The number of words in the editorials ranged from 381 to 903. All counts taken from the texts were normalized to 500 words per texts. The subjects covered in the texts were as follows:

The Falkland Islands	(one text for each language)
Argentine politics	(one text for each language)
the stock market	(one text for each language)
the Iran-Contra affair	(three texts for each language)
elections	(four texts for each language)

The dates of publication for all the texts were within the months of November and December of 1986.

Procedure: The texts were compared on the basis of two types of variables: lexico-syntactic measures and cohesion measures. The specific measures are given in Figure 2 (cf. Figure 1). There were 14 lexico-syntactic measures:

Prepositions (e.g., at, for, on, with; com, no, de, para)
Nominalizations (e.g., -ism, -ity, -ment; -ismo, -ao, -ia)

First and Second Pronouns (e.g., I, you, we; eu, tu)
Third Person Pronouns (e.g., he, she, they; ele, ela, eles)
Precise Conjuncts (e.g., that is, consequently; igualmente, enfim)
General Hedges (e.g., almost, mostly; em geral, de certo modo)
Subordinators (e.g., although, because; assim que, porque)
Locative Adverbs (e.g., after, always, now; acima, agora, fora)
Singular Pro-Verb *DO* (e.g., so do you; tambem, nem)
Past Tense (simple and progressive forms)
Present Tense (simple and progressive forms)
Infinitives ("*to* verb" infinitives; *-er* verb suffix)
That clauses (e.g., I think that...; Eu penso que...)
Words per Sentence

The one cohesion measure was frequency of *repetition*. This measure examined the amount of lexical repetition occurring in a text after removing the 200 most common lexical words (no functional class words were counted). Thus, only words which were content oriented (open class) and which were not the most frequent were counted. This count follows from arguments that lexical repetition is critical to textual coherence (e.g., Hoey 1991, Phillips 1985), that content-based repetition represents a more accurate notion of the lexical repetition which contributes to cohesion (Halliday and Hasan 1976:290), that repetition in science and technical writing is seen as more exact and abstract, and that the increasing frequency of this type of repetition is seen as a measure of writing development (Stotsky 1983).

The choice of these 15 variables for the text analysis was determined by 1) the selection of strong variables in the Grabe model (Figure 1), 2) the ability to perform hand counts, and 3) the ability to create similar linguistic measures for both the English and the Portuguese texts. Figure 2 describes the three factor dimensions which are explored through the 15 linguistic features noted above. The linguistic features do not represent all 31 features used in the original model. It is assumed here that the model is a stable representation of English expository prose and that the features measured would be applied directly to the factor dimensions on which they had appeared in the original model. The stability of the dimensions across factor rotations in the original model as well as the similarities to Biber's modelling provide sufficient support for this assumption for English. The related assumption that this model would be a valid model of Portuguese written prose is yet to be demonstrated, though our second study will indicate that this assumption is plausible.

The counts of the linguistic features in this study required a number of decisions since certain of the measures in English do not equate exactly with the Portuguese counterparts. The counts which depend on specific lexical forms were determined for English by referring to Quirk et al. (1985), and Leech and Svartvik (1977). Comparative categories for the Portuguese data were a matter of interpretation on the part of Dantas-Whitney, working with available Portuguese linguistic resource materials (see Dantas-Whitney and Grabe to appear). These interpretations were judged to be reasonable by a group of Brazilian Applied Linguistics faculty at Catholic University - Rio, who read the Dantas-Whitney and Grabe (to appear) paper as part of a seminar in written discourse analysis.

Factor 1	Factor 2	Factor 3
non-narrative vs. narrative	*orientation of discourse*	*information type*
immediate (+)	*interactional (+)*	*abstract/logical (+)*
present tense	general hedges	nominalizations prepositions
precise conjuncts	1st./2nd. person pronouns infinitives that-clauses subordinators PRO-verb DO	words per sentence
distant (-)	*information (-)*	*situational (-)*
past tense	repetition	3rd. person singular pronoun locative adverbs

Figure 2. Categorization of syntactic and cohesion variables according to three factors (cf. grabe 1987)

The comparative analysis of tenses and pronouns were relatively straightforward. Tenses were determined by verbal suffixes in both languages. Pronouns were assumed to be present in the Portuguese texts when so indicated by verbal morphology. A number of the linguistic variables used in this study were counts based on a specific set of lexical forms in both English and Portuguese (e.g., locative adverbs; Dantas-Whitney and Grabe to appear, for details). The specific occurrence of *Do* as a pro-verb has no direct equivalent in Portuguese; instead, the forms *nem* and *tambem* were taken as equivalents where appropriate to do so. The three most problematic counts for comparison involved two syntactic structure measures, *infinitives* and *that clauses*, and the lexical category of *preposition* (see Appendix B). It is apparent that comparative syntactic counts of English and Portuguese *that clauses* and *infinitives* will create some problems. Two factors, however, mitigate the above set of problems for this study. First, a number of these examples are preferable for spoken Portuguese and may not occur in editorial texts in any great number. Second, in this study *that clauses* and *infinitives* are counted together on factor two, so any imbalance of comparison between *that clauses* and *infinitives* disappears. It is also important to point out that the extent of this problem can be determined in part by the degree to which the syntactic structure counts behave as might be expected.

The question of comparison with prepositions is also problematic. It is clear that Portuguese requires phrasings with prepositions which are not required for English. It may be the case that there are phrasings in English which require prepositions not required in Portuguese, but none come to mind. For this reason, it is perhaps preferable to view Portuguese (and also Spanish) as a language making more extensive use of prepositions for syntactic purposes. Again, it is not clear to what extent the types of

prepositional structures noted above would be prevalent in editorial writing, though it is assumed for this study that the impact will not be great. The quantified results of the prepositional counts, and the resulting interpretations will also take this prepositional imbalance into account.

Data Analysis: The lexico-syntatic and cohesion measures described above were hand counted throughout the 20-text corpus. The resulting raw scores were then standardized, and z-scores were derived for each text for each variable. The variables were then clustered into the three different factors (Figure 2) created by Grabe (1987). The positive syntactic measures and the negative syntactic measures were then added together to derive factor scores for each text. For example, on Factor 1, our first text might have 30 *present tense* markings, 5 *precise conjuncts*, and 15 *past tense* markings. Adding these together (30 + 5 - 15 = 20) creates a Factor 1 score of 20 for the first text. This scoring procedure is repeated for the counts representing Factors 2 and 3. The first text would then have a factor dimension score for each factor. This procedure would then be repeated for the other nineteen texts. When this adding procedure was completed for all the texts, the Portuguese texts were combined to get an average score for all ten texts on Factor 1, on Factor 2, and on Factor 3. The same final procedure was done for the English texts as well. This averaging of the English and Portuguese texts on each factor dimension was then subjected to a t-test to test for group differences.

3.3. *Results:* Table 1 provides the averaged factor scores for both the English and the Portuguese texts. As indicated in Table 1, there was no significant difference between the English and Portuguese texts on Factor 1, "Non-narrative versus Narrative Context." Similarly, Factor 2, the "Orientation to Discourse," did not yield a statistically significant difference, although a trend is discernable. Factor 3, the "Presentation of Information," did indicate a significant difference between English and Portuguese editorial texts.

Table 1
GROUP MEANS AND T-TESTS OF LINGUISTIC VARIABLES ON THREE FACTORS

Groups	Mean z-scores Factor 1	Factor 2	Factor 3
English	- 0.11	- 0.84	1.82
Portuguese	0.24	0.62	- 1.83
t-Ratio (df = 18)	- 0.47	- 1.34	3.44*

*(t-ratio of 2.101 is significant at .05)

3.4. *Discussion:* The results of the t-tests indicate that Factor 1 did not differentiate English and Portuguese editorials. Given the basic genre orientation of this factor, one might expect that a corpus of only editorial texts would not vary strongly on this factor. This result supports the findings of Grabe (1987), which also indicated that editorial texts did not vary greatly on this factor. It also points out that neither set of texts uses extensive past narrative in their editorializing.

The results of the t-test for Factor 2 require some discussion. There is a trend in the means to differentiate the English and Portuguese texts, and perhaps a larger sample would create a significant difference. Factor 2 is an important indicator of interactive features in written genres (cf. Tannen's [1989] discussion of involvement). That is, Factor 2 may be a strong indicator of reader-writer relations. Biber (1986:394) argues that the positive features of his interactive-informational dimension, such as *general hedges, infinitives, that clauses, abverbial subordinators,* and *pro-verb Do,* as being able to "mark a highly inexplicit presentation of meaning —a style in which a single expression can stand for any of several thoughts." Such ambiguous linguistic structuring is typical of language processing under real-time constraints (cf. Sa'Adeddin 1988). Features such as *first* and *second person pronouns* "mark a high degree of interpersonal involvement." In contrast, the negative feature on this factor, *repetition,* marks a highly explicit information content and indicates little interaction.

To the extent that Factor 2 indicates a trend, it is possible to suggest that Portuguese editorial texts are more interactive, involving the reader to a greater extent than do the English editorials. (See also Sa'Adeddin 1988 and Zellermayer 1988.) This difference matches the intuitions for the native Portuguese speaker in this study (Dantas-Whitney). If the behavior of English and Portuguese texts were hypothesized for a larger study according to this notion, one might expect English texts to be more informationally oriented and Portuguese texts to be more interactionally oriented. The results of the means for Factor 2 are suggestive, though they were not statistically significant.

The results of the t-test on Factor 3 reveals a statistically significant difference between English and Portuguese texts. This factor, defining "Presentation of Information," indicates that English and Portuguese editorials differ in the extent to which information is presented abstractly or concretely (situationally; e.g., marked by temporal and locative adverbs). The variables with positive scores "share a function which marks highly abstract nominal content and a highly learned style" (Biber 1986:395). The variables with negative scores (*third person pronouns* and *locative adverbs*) "share the marking of very concrete content and more informal style" (Biber 1986:395). The differences between the English and the Portuguese texts in relation to Factor 3 suggest that English editorials present clear characteristics of a formal, detached style. Portuguese editorials are significantly more concrete and colloquial (cf. Sa'Adeddin 1988 and Zellermayer 1988). These results are particularly striking since the potential imbalance of prepositional phrases towards Portuguese should have diminished the differences in the means of the Factor 3 scores. The fact that Portuguese syntactic prepositions did not obscure a significant difference indicates that the specific syntactic uses of prepositions in Portuguese, over and above the functional roles they play in both languages, may not be significant for contrastive rhetoric studies.

Through an examination of English and Portuguese editorials, this study has identified at least one significant distinction between the two languages among the three dimensions examined. The significant difference on Factor 3 between English and Portuguese texts indicates that English editorials are more formal while Portuguese editorials are more concrete and colloquial. The trend found in Factor 2, although not statistically significant, is compatible with the text comparison on Factor 3; that is, texts which are more colloquial and concrete seem also to be more interpersonally oriented than informationally oriented (Biber 1988, Grabe 1987).

This exploratory study has also pointed out a number of difficulties with contrastive rhetoric research comparing two languages, particularly with the problem of syntactic equivalences. It illustrates the need for a more careful study of specific linguistic comparisons between English and Portuguese. Factor 3 also demonstrates that a simple listing of potential differences between languages is not sufficient (e.g., different uses of prepositions in two languages). Studies must be done which indicate the frequency with which these differences actually arise on a regular basis, and in what types of texts. This study also points out the need to perform multivariate analyses on large sets of Portuguese texts to determine whether or not a model of factor dimensions for Portuguese texts would look similar to the textual dimensions for English proposed by Biber (1988) and Grabe (1987). The fact that the results of this exploratory study are readily interpretable is suggestive of similarities in textual dimensions across languages. The second study reported here, based on Lux (1991), will provide further support for the cross-language stability of the model used in this study.

4. Comparing English and Spanish University Student Writing

4.1. *Background:* In this second study, groups of students with similar academic backgrounds were asked to write on a topic in either English or Spanish. A major problem with prior research on student writing in contrastive rhetoric studies has been that the cultural factor and the linguistic factor could not be examined separately. The design of this study, however, makes it possible to examine cultural influences (comparing the Ecuadorian Spanish-speaking writers and the US English-speaking writers) independently from language influences (some of each group writing in Spanish or in English). This study was designed also to provide a further test of multivariate approaches to research on contrastive rhetoric. Building on the studies by Biber (1988), Grabe (1987) and Dantas-Whitney and Grabe (to appear), this study counts syntactic and cohesion variables assumed to indicate differences in expository prose, particularly in student writing. These variables are then combined into a factor analysis which creates a distinct factor dimensional structure for cross-language student writing. As will be seen, the similarities with the expository prose model for English (Grabe 1987) are far greater than the differences, despite the introduction of a new language (Spanish), a new culture (Ecuadorian university students), and possibly a new genre (student writing, perhaps a sub-genre of expository prose). Finally, this study further demonstrates the applicability of a multivariate text analytic approach to contrastive rhetoric.

4.2. *Method:* This study of Ecuadorian and US university student writing was designed specifically to measure the influence of culture, language, and second language development on 20 text features which have been identified (Grabe 1987, Ostler 1990, Reid 1988) as having interpretable discourse functions.

Subjects and writing assignment: The subjects were 158 post-secondary students from Ecuador and the United States who were asked by their teacher to write on an assigned topic in their regular college classes. The writing prompt, adapted from a study by Connor (1984), asked the students to argue for or against the value of testing in college-level courses. This topic, while academically oriented, was sufficiently general

so that student writers did not require special background knowledge. The topic was also seen as one with minimal cultural-content bias. Student writers were given the topic on a handout one class before the class session in which they wrote. This was done in order to give students a chance to do their best. The actual writing was done without notes, however, and the students were given one class period to complete their essay (approximately 50 minutes).

The writers were of two cultural backgrounds: native English-speaking Anglo-Americans from the U.S. and native Spanish-speaking Latin Americans from Ecuador. A sub-group within each of these culturally defined groups wrote in English while a second sub-group wrote in Spanish. Thus, four distinct groups were identified for comparison in this study:

a) Anglo-Americans writing in English (AE, n = 41)
b) Anglo-Americans writing in Spanish (AS, n = 36)
c) Latin Americans writing in English (LE, n = 30)
d) Latin Americans writing in Spanish (LS, n = 51)

The mean age of all the writers was 20.19. Although the English L1 group was somewhat younger on average (19.17), comparative ANOVA measures indicated that there were no significant group differences in age or gender mix.

The 158 texts used in this study were drawn from a larger set of approximately 250 essays on the basis of a number of controlling criteria, including writer's age (16-26), extent of post-secondary instruction in writing (none), and overseas exposure to other cultures (less than one year). The latter two criteria were established to diminish the role of formal education as a cultural influence (Reid 1988) and to eliminate writers who were bicultural to a significant degree. Most of the L1 writers were in the first weeks of a required academic writing course; however, some of the Ecuadorian writers were in the first weeks of an introductory content-area course. All of the L2 writers were enrolled in an advanced-level second-language course for which a writing assignment like the one given would have been appropriate. Unfortunately, the criterion of "no previous writing class" had to be relaxed for the group of Anglo-Americans who wrote in Spanish since these writers had typically been enrolled at the university for one or more semesters already, during which time they would have been required to take Freshman English.

Procedure: The comparisons of student texts were based on 17 of the 33 syntactic and cohesion counts found in Grabe (1987) plus three additional text variables: *word length, essay length,* and *coordinated clauses.* These latter three measures have been used in other current research comparing Spanish and English texts (Ostler 1990, Reid 1988). The text variables taken from the Grabe (1987) study were those most easily translatable from English to Spanish. In all, the 20 text features used in this study were:

Prepositions (e.g., for, with, to; por, en, con)
Nominalizations (e.g., -ity, -ism, -ment; -ción, -miento)
Adverbial Subordinators (e.g., because, since; porque, ya que)
First and Second Pronouns (e.g., I, you, we; yo, usted, nosotros)
Third Person Pronouns (e.g., he, she, they; él, ella, ellos)
Locative Adverbs (e.g., now, after; ayer, después)

Conjunctive Adverbs (e.g., then, in addition; además, sin embargo)
Passive Voice (e.g., was rejected; es considerada, se habla mucho de)
WH Nominal Clauses (e.g., I know who lost it; no expresa lo que realmente sabe)
That Nominal Clauses (e.g., I know that he left; pienso que es algo voluntario)
Non-finite Nominal Clauses (e.g., I expect him to be fair; nos permite desarrollar nuestro potencial)
Adjective Clauses (e.g., exams which are given weekly; la atención que debemos poner al profesor)
Questions (e.g., What does this test measure?; ¿Qué podemos hacer?)
Coordinated Clauses (e.g., Some students have no trouble, and they don't feel any pressure; Los estudiantes tenemos exámenes, y siempre hay problemas)
Past Tense (same as Study 1)
Present Tense (same as Study 1)
Repetition (same as Study 1)
Word Length
Words per Sentence
Essay Length

Despite the fact that these text features were considered the most readily translatable from English to Spanish, problems in measuring the text features cross-linguistically were considerable. First, there was the problem in identifying equivalent features in the two languages (similar to Study 1 above). As recognized by Dantas-Whitney and Grabe (to appear), the feature *first* and *second person pronouns* had to be generalized to include both overt pronouns as well as verb endings indicating first and second pronouns for pronoun-dropping languages such as Portuguese or Spanish. Second, in cases in which a text feature is counted because it is on a finite list (e.g., *locative adverbs, prepositions, nominalizations*), problems of equivalence arise in compiling these lists. This problem becomes more complex when cognate forms have different functional interpretations in various discourse uses. The problem of equivalence lists, however, can be diminished by consulting a number of native-speaker linguists in each language and by reading a number of texts similar to the ones in question as a test of the lists.

A third problem which arises in such comparisons across languages is with the assumption that there is equivalence in communicative functions (in both languages) for each feature. While it may be reasonable to assume, for example, that *first* and *second person pronouns* contribute to a more "interactive" text in both Spanish and English, it is not as clear that *that clauses* in Spanish fulfill the "interactive" function for English identified in Grabe (1987) and Biber (1988). This is a problem that was left unresolved in our first study.

In response to the first two problems noted above, this second study attempted to refine a number of the criteria for equivalence proposed in Dantas-Whitney and Grabe (to appear). In response to the third problem, the present study performed a separate factor analysis of the linguistic features examined in the student texts. This factor analysis created factor dimensions only from the data of the student texts, and did not depend on a set of assumed textual dimensions generated from an English corpus of texts. In this way, it was expected that groups of variables would more accurately represent potential intercultural dimensions. At the same time, this approach provides

a reasonable test of the English factor structure assumed in our first study.

This second study also encountered problems which were not issues in study one. Because this study examined student writing, and some of the writing was in a second language, there was a writing proficiency issue which had to be considered. For example, a second-language writer may be writing in a context which clearly requires past tense verbs, but may put only half of these verbs in past tense form. While recognizing these inconsistencies, this study as a rule identified and counted the forms which actually appeared, whether their appearance was appropriate or not. This decision kept the counting procedure consistent across all texts and avoided subjective interpretations of the students' writing. The potential for this issue to be a confounding variable is recognized and is considered in the discussion of results.

Data Analysis: After all variables in the twenty texts were counted, a factor analysis of the variables was performed. Groups of variables which co-occur in the texts will combine to form factors, which are interpreted as textual dimensions. In this way, factor analysis reduces the large number of individual linguistic features to a more manageable number of factors. In the process, it is expected that these factors will better describe the text corpus by showing functional connections between text features which may not be apparent from the analysis of the features individually.

Table 2
FACTOR STRUCTURE: FACTOR LOADINGS OF 20 TEXT VARIABLES

Variable	Factor 1	Factor 2	Factor 3	Factor 4
PREPS.	.67392 *	-.07371	-.42614 *	-.20466
NOMINALIZA.	.32138	.16576	-.49553 *	-.05790
1 & 2 PRNS.	-.19309	-.53161 *	.29946	-.13389
3 PRNS.	-.30000	-.14677	.10007	.01840
PAST TENSE	-.05083	.03370	.28360	-.58375 *
PRES. TENSE	-.61038 *	-.47565 *	.06543	.71600 *
WORDS/T-UNIT	.89238 *	-.10169	-.17739	.05051
CONJUNCT. ADV.	.06817	.26090	.10219	-.05273
SUBORDINATORS	-.13219	-.29779	.11558	.07258
PASSIVES	.26530	.25100	-.11723	-.02433
WH-CLAUSES	-.22178	.01065	.44761 *	.20922
THAT-CLAUSES	-.04182	-.51306 *	.07855	.30492
NOMINAL PHR.	.19856	.23889	.11879	-.09182
ADJ. CLAUSES	.45430 *	.02380	-.06427	-.02361
QUESTIONS	-.09290	.08399	.21440	-.08096
LOCATIVES	-.19421	.27721	.12780	.05543
REPETITION	-.23646	.56925 *	.22364	.11751
WORD LENGTH	.04920	.19665	-.60208 *	.13066
ESSAY LENGTH	.13719	.28287	.34501	.14555
COORD. CLAUSES	-.39257 *	-.15673	.02913	-.39990 *

Note. Substantial loadings (> .35) are marked with an asterisk.

A four-factor solution was determined to be the best one for the data (Lux 1991). The four factors, which together explained 32% of the variance of the 20 linguistic variables, are shown in Table 2. Assuming a .35 cutoff value, we see that five linguistic variables load on Factor 1: *Prepositions, words per sentence,* and *adjective clauses* with

positive values; *present tense* and *coordinated clauses* with negative values. Each of the factors is characterized in this way, with Factors 2, 3, and 4 having four, four, and three substantial loadings, respectively (note asterisks in Table 2).[1]

Based upon the combinations of linguistic variables loading on the factors, each factor was given an interpretive label.[2] In the present case, characterizing the text dimensions was aided by the numerous correspondences that were found between this factor structure and that of Grabe (1987). For example, Factor 2 here includes *first* and *second pronouns, present tense,* and *that clauses* on the negative side, and *word repetition* on the positive side. All of these variables appear with the same opposition in Grabe's (1987) Factor 2, which he labels "Interactional vs. Informational Orientation" (see Figure 1). Factors 3 and 4 also have correspondences to Grabe's (1987) study. Factor 3 in this study has much in common with Grabe's (1987) Factor 3, which he labels "Abstract vs. Situational Information Type." Factor 4 here, despite its rather meager loadings, is quite clearly a narrative/non-narrative dimension corresponding to Grabe's Factor 1, "Immediate vs. Distant Context."

The most distinctive factor evolving out of the present corpus, and the strongest factor in this study, is Factor 1. Factor 1 is characterized by *long sentences, frequent prepositions,* and *frequent adjectival clauses* (relative clauses), on the one hand, and frequent *present tense* and *coordinated clauses,* on the other. It seems to represent a sentence-level elaboration dimension in the texts, with some texts having sentences lengthened by additional arguments being placed in subordinate structures, and other texts having these additional arguments in new independent structures. This dimension has been labelled "Elaborated vs. Reduced Sentence Style" following recent work by Biber (1990) on syntactic complexity.

In a manner similar to our first study, the linguistic features on each factor were added to create factor scores for each text. In this way, it is possible to compare the patterns of variation created by each group of student writers (AE, AS, LE, LS) along each textual dimension. The creation of these factor-score means permits the factor analysis to be tested to see whether or not it captures significant differences among the groups of writers. For this purpose, a multiple analysis of variance (MANOVA) was performed. Based on the significance of the MANOVA, ANOVAs for each factor were examined and Scheffé tests were performed.

4.3. *Results:* The overall significance value for the MANOVA (.000) indicated that

---

[1] The factor analysis used was a Principal Factors approach with an oblique rotation. The oblique rotation assumes some degree of interdependence among the factors generated, which is a plausible assumption for complex linguistic data. The factor structure created by the procedure tells how each variable "loads" on (or contributes to) each factor. On the basis of suggestions in Gorsuch (1983), a factor coefficient of .35 was considered substantial for this 158 text corpus. (A negative value indicates an inverse relation to other features on the factor.) The viability of a factor is measured both through the number of factor loadings and through Eigenvalues, which indicate the amount of variance in the corpus explained by the factor. Final statistics for the factor analysis showed that Eigenvalues for Factors 1-4 in this study were 2.77, 1.66, 1.27, and .896, respectively.

[2] Factor interpretation is a challenging and subtle aspect of factor analysis, and the subjective nature of such interpretations is validated through the explanatory power offered by the labelling as well as by supporting research.

the writing groups performed significantly differently with respect to the four factors generated from the corpus. The component ANOVAs indicated that the independent variables 'Culture' and 'Language of Task,' as well as the interaction effects of these two variables, distinguished the texts at an F significance value of .000. Table 3 shows the strength for Culture, Language of Task, and the Culture/Language interaction as predictors of variance on each of the four factors. In order to explore in more detail the contributions of each variable to the patterns of variation among groups of writers, Scheffé tests were performed for each component ANOVA. Results of the Scheffé tests are presented in Table 4.

Table 3
PREDICTORS OF VARIATION ON FACTORS 1-4: COMPONENT ANOVA RESULTS

Factor Variable	Source of Variation	F Ratio	F Probability (Significance)
Factor 1	Culture	62.8864	.000*
Reduced/	Language	2.0187	.157
Elaborated	Culture/Lang.	68.0077	.000*
Factor 2	Culture	3.7406	.055
Interaction/	Language	27.2115	.000*
Information	Culture/Lang.	12.0530	.000*
Factor 3	Culture	49.4944	.000*
Abstract/	Language	14.9039	.000*
Conversational	Culture/Lang.	26.2177	.000*
Factor 4	Culture	3.4074	.067.
Narrative/	Language	4.3266	.039
Non-Narr.	Culture/Lang.	23.8062	.000*

Note. Asterisk indicates a statistical significance level of $p < .01$. Non-significant interactions are not shown.

Table 4
GROUP DIFFERENCES ON FACTORS 1-4: SCHEFFÉ TESTS

Factor Variable	Writing Group/Mean		Significant Group Differences ($p < .05$)		
Factor 1	AS	-.0203	LE	>	AS
Reduced (-)	LE >	16.7554 *	AE	>	AS
Elaborated (+)	AE >	17.4524 *	LS	>	AE, LE, AS
	LS	39.6694 ***			
Factor 2	AS	-10.6028	AE	>	LE, LS, AS
Interaction (-)	LS	-7.2322			
Information (*)	LE	-3.8446			
	AE	15.4330 ***			

Factor 3	LS	-42.4671	AS	>	LS
Abstract (-)	LE	-30.9926 *	AE	>	LS
Conversational (+)	AE	-27.3202 *	LE	>	LS
	AS	-25.2669 *			
Factor 4	AE	23.1064	AS	>	LS, AE
Narrative (-)	LS	24.0715			
Non-	LE	28.9710			
narrative (*)	AS	35.2398 **			

Note. Asterisk (*) means that this group was significantly different from another group ($p < .05$) -two asterisks, two groups, etc. AE = Anglo-Americans writing in English; AS = Anglo-Americans writing in Spanish; LE = Latin Americans writing in English; LS = Latin Americans writing in Spanish. The notation AS > AE indicates that AS group mean was significantly higher than that of the AE group.

4.4. *Discussion:* On the basis of the group mean scores for Factor 1 and the results of the Scheffé test, it appears that variation on Factor 1, "Elaborated vs. Reduced Sentence Style," can be attributed to both cultural and developmental influences. Figure 3 presents the differences in group means for Factor 1. As indicated by Figure 3, the sentences of the Latin-American writers (both L1 and L2) are much more syntactically elaborated than those of the corresponding (L1 and L2) Anglo-American writers, suggesting a Latin American preference for syntactic elaboration. This finding supports results from other contrastive research on student writing (Montano-Harmon 1988, 1991, Ostler 1987, 1990, Reid 1988). At the same time, the fact that both groups of L2 writers employed a more "reduced" sentence style, compared to their same-culture counterparts writing in the L1, indicates that developmental factors clearly play a role in sentence style as well. These two findings are also indicated in the Scheffé results; both the Latin American Spanish writers and the Anglo-American Spanish writers are significantly different from all other groups.

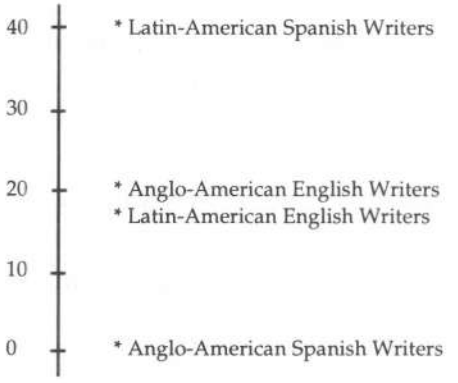

Figure 3: Mean group scores on Factor 1:
"Elaborated vs. Reduced Sentence Style"

A review of the mean scores on Factors 2 and 3 show that the texts of the Anglo L1 writers were determined to be much less interactional than those of the other groups and that the texts of the Latin American L1 writers were found to be much more

abstract and formal (see Factor Score Means for Factors 2, 3, and 4 in Appendix C). These apparent cultural tendencies appear to neutralize, however, when the Anglo and Latin writers performed in the L2; this would suggest that the ability to signal a strong Informational Orientation (for English writers) or an Abstract Informational Style (for Spanish writers) develops in advanced levels of L1 writing proficiency.

Factor 4, which in the present corpus seems to represent a tendency for the writer to drop into narrative, differentiates the four groups of writers the least. A lack of cultural difference is not surprising on the basis of our first study, which found Brazilian Portuguese and American English editorials to be quite similar on a "Narrative vs. Non-narrative" dimension. Since this dimension is most likely a genre-based dimension of textual variation, it is not surprising that a corpus comprised essentially of a single genre would not indicate culture or language-based variation. The fact that this factor was strong in Grabe (1987), as Factor 1, and relatively weak in this study, as Factor 4, highlights the difference in genre emphasis the two corpora used. Factor 4 does indicate a developmental variable at work, however. For example, both second-language groups show a tendency toward the non-narrative pole while L1 writers seem to be more willing to include narrative elements in their essays.

To summarize, this study has generated several textual dimensions which are valid for cross-language comparisons. Based on this analysis, we can say that Latin-American Spanish writers prefer an elaborated sentence style in their writing, and they appear to prefer a more abstract informational presentation in their writing. In contrast, Anglo-American English writers appear to prefer a more reduced sentence style, as well as a more informationally oriented (and less interactional) style. Factor 4 did not indicate any clear cultural preferences, as might be expected.

These results, particularly when disregarding the L2 writers, generally agree with the first study, of Portuguese and English editorials. Both found that the genre-based factor, "Narrative vs. Non-narrative," did not indicate significant group differences. They also agreed in the view that English is more informationally oriented and less interactional. The "Elaborated vs. Reduced Sentence Style" factor appears to be a true cross-linguistic dimension in that it did not appear in Grabe's (1987) study of English edited prose text types, and that it matches previous findings in contrastive rhetoric research (Montano-Harmon 1988, 1991). The results of group variation on Factor 3, appear to create some contradiction with the results of the first study in that English editorials appeared to be more abstract and less situational than the Portuguese editorials, but the Spanish essays were more abstract and less conversational than the English essays. This contradiction may indicate a difference between Spanish and Portuguese prose, or a difference between editorial writing and student essay writing; or it may suggest that larger corpora should be created and more linguistic features examined.

5. Conclusion

This second study has applied Biber's multivariate approach directly to a corpus of cross-linguistic student writing. In doing so, it has demonstrated the applicability of such a methodology to contrastive rhetoric research, thus confirming Grabe's (1987) argument for the utility of such an application. The second study also confirms the basic factor structure analysis presented in Grabe (1987) and used by Dantas-Whitney

and Grabe (to appear) in their exploratory study. The two studies together present a strong argument for the use of Biber's multivariate approach to text analysis. At the same time, the two studies using the approach reported here support many other findings in contrastive rhetoric research (e.g., Hinds 1987, Jenkins and Hinds 1987, Montano-Harmon 1988, 1991, Ostler 1987, 1990, Reid 1988, Sa'Adeddin 1988, Zellermayer 1988). This multivariate approach to text analysis —interpreting co-occurrences of linguistic features in texts as functional dimensions of written discourse— allows for more powerful analyses than most other text-based research. We believe that such multivariate analyses of cross-language text corpora will continue to provide important insights for the notion of contrastive rhetoric.

REFERENCES

ATKINSON, D. (1991). Discourse analysis and written discourse conventions. *Annual Review of Applied Linguistics, 1990* 11: 57-76.

BAZERMAN, C. (1988). *Shaping written knowledge*. Madison, WI: University of Wisconsin Press.

BAZERMAN, C. (1991). Discourse analysis and social construction. *Annual Review of Applied Linguistics, 1990* 11: 77-83.

BERMAN, R. (1985). The acquisition of Hebrew. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition. Volume 1: The data*. Hillsdale, NJ: L. Erlbaum Associates. 255-371.

BESNIER, N. (1988). The linguistic relationships of spoken and written Nukulaelae registers. *Language* 64: 707-736.

BIBER, D. (1985). Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. *Linguistics* 23: 337-360.

BIBER, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62: 384-414.

BIBER, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.

BIBER, D. (1991). On the complexity of discourse complexity: A multi-dimensional analysis. Unpublished ms., Northern Arizona University.

BICKNER, R. and P. PEYASANTIWONG. (1988). Cultural variation in reflective writing. In A. Purves (Ed.), *Writing across languages and cultures*. Newbury Park, CA: Sage, 160-174.

BIRCH, D. (1989). *Language, literature and critical practice*. London: Routledge.

BOGGS, S. (1985). *Literacy, learning and schooling*. Norwood, NJ: Ablex.

BRUFFEE, K. (1986). Social construction, language, and the authority of knowledge: A bibliographic essay. *College English* 48: 773-790.

CARLSON, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures*. Newbury Park, CA: Sage. 227-260.

CAZDEN, C. (1988). *Classroom discourse*. Portsmouth, NH: Heinemann.

CERNIGLIA, C., K. MEDSKER and U. CONNOR. (1990). Improving coherence by using computer-assisted instruction. In U. Connor and A. Johns (Eds.), *Coherence in writing*. Washington, DC: TESOL Publications. 227-241.

CHAFE, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language*. Norwood, NJ: Ablex. 35-54.

CHENG, P. (1985). An analysis of contrastive rhetoric: English and Chinese expository prose, pedagogical implications, and strategies for the ESL teacher in a ninth-grade curriculum. *Dissertation Abstracts International* 46: 2537A.

CLANCY, P. (1986). The acquisition of communicative style in Japanese. In B. Schieffelin and E. Ochs (Eds.), *Language socialization across cultures*. New York: Cambridge University Press. 213-250.

CLYNE, M. (1983). Linguistics and written discourse in particular languages: Contrastive studies: English and German. *Annual Review of Applied Linguistics, 1982* 3: 38-49.

CLYNE, M. (1985). *Language and society in the German-speaking countries.* New York: Cambridge University Press.
CLYNE, M. (1987). Cultural differences in the organization of academic texts. *Journal of Pragmatics* 11: 211-247.
COLLINS, J. and S. MICHAELS. (1986). Speaking and writing: Discourse strategies and the acquisition of literacy. In J. Cook-Gumperz (Ed.), *The social construction of literacy.* New York: Cambridge University Press. 207-222.
CONNOR, U. (1984). Recall of text: Differences between first and second language readers. *TESOL Quarterly* 18: 239-255.
CONNOR, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the teaching of English* 24: 67-87.
CONNOR, U. and J. LAUER. (1988). Cross-cultural variation in persuasive student writing. In A. Purves (Ed.), *Writing across languages and cultures.* Newbury Park, CA: Sage. 138-159.
CONNOR, U. and A. JOHNS (Eds.). (1990). *Coherence in writing.* Washington, DC: TESOL Publications.
CROWLEY, S. (1990). *A teacher's introduction to deconstruction.* Urbana, IL: National Council of Teachers of English.
DANTAS-WHITNEY, M. and W. GRABE. (to appear). A comparison of Portuguese and English newspaper editorials. In T. Ricento (Ed.), *Research on contrastive rhetoric.* Washington, DC: TESOL Publications.
DERRIDA, J. (1976). *Of grammatology.* Baltimore, MD: John Hopkins University Press.
EGGINGTON, W. (1987). Written academic discourse in Korean: Implications for effective communication. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text.* Reading, MA: Addison-Wesley. 153-168.
ERVIN-TRIPP, S. (1972). On sociolinguistic rules: Alternation and co-occurrence. In J. Gumperz and D. Hymes (Eds.), *Directions in sociolinguistics.* New York: Holt, Rinehart and Winston. 213-250.
EVENSON, L. (1990). Pointers to superstructure in student writing. In U. Connor and A. Johns (Eds.), *Coherence in writing.* Washington, DC: TESOL Publications. 169-183.
FAIRCLOUGH, N. (1989). *Language and power.* London: Longman.
GORSUCH, R. (1983). *Factor analysis.* 2nd Ed. Hillsdale, NJ: Erlbaum Associates.
GRABE, W. (1987). Contrastive rhetoric and text type research. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 texts.* Reading, MA: Addison-Wesley. 115-138.
GRABE, W. (1990). Current developments in written discourse analysis. *Lenguas Modernas* 17: 35-56.
GRABE, W. and R.B. KAPLAN. (1989). Writing in a second language: Contrastive rhetoric. In D. Johnson and D. Roen (Eds.), *Richness in writing: Empowering ESL students.* New York: Longman. 263-283.
HALLIDAY, M. (1985). An introduction to functional grammar. London: Edward Arnold.
HALLIDAY, M. and R. HASAN. (1976). *Cohesion in English.* London: Longman.
HALLIDAY, M. and R. HASAN. (1989). *Language, context, and text: Aspects of language in social semiotic perspective.* New York: Oxford University Press.
HEATH, S.B. (1983). *Ways with words.* New York: Cambridge University Press.
HEATH, S.B. (1986a). Critical factors in literacy development. In S. deCastell, A. Luke and K. Egan (Eds.), *Literacy, society, and schooling: A reader.* New York: Cambridge University Press. 209-229.
HEATH, S.B. (1986b). Sociocultural contexts of language development. In California State Department of Education (Ed.), *Beyond language: Social and cultural factors in schooling language minority students.* Los Angeles: California State University. Evaluation, Dissemination and Assessment Center. 143-186.
HINDS, J. (1983). Linguistics and written discourse in particular languages: Contrastive studies: English and Japanese. *Annual Review of Applied Linguistics, 1982* 3: 78-84.
HINDS, J. (1987). Reader versus writer responsibility: A new typology. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text.* Reading, MA: Addison-Wesley. 141-152.
HINDS, J. (1990). Inductive, deductive, quasi-inductive: Expository writing in Japanese, Korean, Chinese, and Thai. In U. Connor and A. Johns (Eds.), *Coherence in writing.* Washington, DC: TESOL Publications. 89-109.
HOEY. M. (1991). *Patterns of lexis in text.* London: Oxford University Press.
HUNT, E. and F. AGNOLI. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review* 98: 377-389.
HYMES, D. (1974). *Foundations of sociolinguistics.* Philadelphia, PA: University of Pennsylvania Press.
JENKINS, S. and J. HINDS. (1987). Business letter writing: English, French, and Japanese. *TESOL Quarterly* 21: 327-345.
INDRASUTA, C. (1988). Narrative styles in the writing of Thai and American students. In A. Purves (Ed.), *Writing across languages and cultures.* Newbury Park, CA: Sage. 206-226.

KACHRU, Y. (1983). Linguistics and written discourse in particular languages: Contrastive studies: English and Hindi. *Annual Review of Applied Linguistics, 1982* 3: 50-77.

KACHRU, Y. (1988). Writers in Hindi and English. In A. Purves (Ed.), *Writing across languages and cultures*. Newbury Park, CA: Sage. 109-137.

KAPLAN, R.B. (1966). Cultural thought patterns in intercultural education. *Language Learning* 16: 1-16.

KAPLAN, R.B. (1972). *The anatomy of rhetoric: Prolegomena to a functional theory of rhetoric*. Philadelphia, PA: Center for Curriculum Development.

KAPLAN, R.B. (1987). Cultural thought patterns revisited. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley. 9-21.

KAPLAN, R.B. (1988). Contrastive rhetoric and second language learning: Notes toward a theory of contrastive rhetoric. In A. Purves (Ed.), *Writing across languages and cultures*. Newbury Park, CA: Sage, 275-304.

KAPLAN, R.B. and W. GRABE. (1991). The fiction in science writing. In H. Schroeder (Ed.), *Subject-oriented texts*. New York: Walter deGruyter. 199-217.

KRESS, G. (1989). *Linguistic processes in sociocultural practice*. New York: Oxford University Press.

KRESS, G. (1991). Critical discourse analysis. *Annual Review of Applied Linguistics, 1990* 11: 84-99.

LAUTTAMATTI, L. (1987). Observations on the development of the topic of simplified discourse. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley. 87-114.

LEECH. G. and J. SVARTVIK. (1977). *A communicative grammar of English*. New York: Longman.

LEKI, I. (1991). Twenty-five years of contrastive rhetoric: Text analysis and writing pedagogies. *TESOL Quarterly* 25: 123-143.

LUX, P. (1991). *Discourse styles of Anglo and Latin American college student writers*. Phoenix, AZ: Arizona State University, Ph.D. diss.

MARTIN, J. (1989). *Factual writing*. New York: Oxford University Press.

MONTANO-HARMON, M. (1988). *Discourse features in the compositions of Mexican English as a second language, Mexican-American Chicano, and Anglo high school students: Considerations for the formulation of educational policy*. Los Angeles: University of Southern California, Ph.D. diss.

MONTANO-HARMON. (1991). Discourse features of written Mexican Spanish: Current research in contrastive rhetoric and its implications. *Hispania* 74: 417-425.

MYERS, G. (1985). The social construction of two biologists' proposals. *College English* 48: 595-610.

NEEL, J. (1988). *Plato, Derrida, and writing*. Carbondale, IL: Southern Illinois University Press.

OCHS, E. (1988). *Culture and language development*. New York: Cambridge University Press.

OSTLER, S. (1987). English in parallels: A comparison of English and Arabic prose. In U. Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley. 169-185.

OSTLER, S. (1990). The contrastive rhetorics of Arabic, English, Japanese, and Spanish. Paper presented at the 25th Annual TESOL Convention, San Francisco, CA, March, 1990.

PHILLIPS, M. (1985). *Aspects of text structure: An investigation of the lexical organization of text*. Amsterdam: North Holland.

PHILLIPS, S. (1983). *The invisible culture*. New York: Longman.

PURVES, A. and G. HAWISHER. (1990). Writers, judges, and text models. In R. Beach and S. Hynds (Eds.), *Developing discourse practices in adolescence and adulthood*. Norwood, NJ: Ablex. 183-199.

QUIRK, R., J. GREENBAUM, G. LEECH and J. SVARTVIK. (1985). *A comprehensive grammar of English*. New York: Longman.

REID, J. (1988). *Quantitative differences in English prose written in Arabic, Chinese, Spanish, and English students*. PhD diss. Colorado State University.

SA'ADEDDIN, M. (1988). Text development and Arabic-English negative Interference. *Applied Linguistics* 9: 36-51.

SCHIEFFELIN, B. and P. GILMORE (Eds.). (1986). *The acquisition of literacy*. Norwood, NJ: Ablex.

SCHIEFFELIN, B. and E. OCHS (Eds.). (1986). *Language socialization across cultures*. New York: Cambridge University Press.

SCOLLON, R. and S. SCOLLON. (1981). *Narrative, literacy and face in interethnic communication*. Norwood, NJ: Ablex.

SHOR, I. (Ed.). (1987). *Freire and the classroom*. Portsmouth, NH: Boynton/Cook.

SLOBIN, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Micaelis, and H. Filip (Eds.), *Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting*. Berkeley, CA: Berkeley Linguistic Society. 435-445.

SLOBIN, D.I. and A. BOCAZ. (1988). Learning to talk about movement through time and space: The development of narrative abilities in Spanish and English. *Lenguas Modernas* 15: 5-24. [Circulated as Berkeley Cognitive Science Report Nº 55, Institute of Cognitive Studies, University of California, Berkeley, January 1989].

SOTER, A. (1988). The second language learner and cultural transfer in narration. In A. Purves (Ed.), *Writing across languages and cultures*. Newbury Park, CA: Sage. 177-205.

STOTSKY, S. (1983). Types of lexical cohesion in expository prose: Implications for developing the vocabulary of academic discourse. *College Composition and Communication* 34: 430-446.

STREET, B. (1984). *Literacy in theory and practice*. New York: Cambridge University Press.

SWALES, J. (1990). *Genre analysis*. New York: Cambridge University Press.

TANNEN, D. (1982). Oral and literate strategies in spoken and written narratives. *Language* 58: 1-21.

TANNEN, D. (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. New York: Cambridge University Press.

WODAK, R. (Ed.). (1989). *Language, power and ideology*. Philadelphia, PA: John Benjamins.

ZELLERMAYER, M. (1988). An analysis of oral and literate texts: Two types of reader-writer relationships in Hebrew and English. In B. Rafoth and D. Rubin (Eds.), *The social construction of writing*. Norwood, NJ: Ablex. 287-303.

APPENDIX A

FACTOR STRUCTURE LOADINGS FROM BIBER'S (1988) STUDY SUMMARY OF THE FACTORIAL STRUCTURE

FACTOR 1

private verbs	.96
THAT deletion	.91
contractions	.90
present tense verbs	.86
2nd person pronouns	.86
DO as pro-verb	.82
analytic negation	.78
demonstrative pronouns	.76
general emphatics	.74
1st person pronouns	74
pronoun IT	.71
BE as main verb	.71
causative subordination	.66
discourse particles	.66
indefinite pronouns	.62
general hedges	.58
amplifiers	.56
sentence relatives	.55
WH questions	.52
possibility modals	.50
non-phrasal coordination	.48
WH clauses	.47
final prepositions	.43
(adverbs	.42)
(conditional subordination	.32)
— — — — — — — — — — — —	—
nouns	-.80
word length	-.58
prepositions	-.54
type/token ratio	-.54
attributive adjs.	-.47
(place adverbials	-.42)
(agentless passives	-.39)
(past participial WHIZ deletions	-.38)
(present participial WHIZ deletions	-.32)

FACTOR 2

past tense verbs	.90
third person pronouns	.73
perfect aspect verbs	.48
public verbs	.43
synthetic negation	.40
present participial clauses	.39
— — — — — — — — — — — —	—
(present tense verbs	-.47)
(attributive adjs.	-.41)
(past participial WHIZ deletions	-.34)
(word length	-.31)

FACTOR 3

WH relative clauses on object positions	.63
pied piping constructions	.61
WH relative clauses on subject positions	.45
phrasal coordination	.36
nominalizations	.36
— — — — — — — — — — — —	—
time adverbials	-.60
place adverbials	-.49
adverbs	-.46

FACTOR 4

infinitives	.76
prediction modals	.54
suasive verbs	.49
conditional subordination	.47
necessity modals	.46
split auxiliares	.44
(possibility modals	.37)
— — — — — — — — — — — —	—
—no negative features —	

FACTOR 5

conjuncts	.48
agentless passives	.43
past participial clauses	.42
BY-passives	.41
past participial WHIZ deletions	.40
other adverbial subordinators	.39
(predicative adjs.	31)
(type/token ratio	-.31

FACTOR 6

THAT clauses as verb complements	.56
demonstratives	.55
THAT relative clauses on object positions	.46
THAT clauses as adj. complements	.36
(final prepositions	.34)
(existential THERE	.32)
(demonstrative pronouns	.31)

(WH relative clauses on object positions	.30)	(downtoners	.33)
(phrasal coordination	-.32)	(adverbs	.31)
		(concessive subordination	.30)
FACTOR 7		(atributive adjs.	.30)
		— — — — — — — — — — — —	
SEEM / APPEAR	.35	—no negative features—	

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

(Adapted from Biber (1988), *Variation across speech and writing.*)

APPENDIX B
PROBLEMS WITH ENGLISH AND PORTUGUESE SYNTACTIC EQUIVALENCES

THAT CLAUSES AND INFINITIVES

There are many cases in which *that clauses* are used in Portuguese, but *infinitives* would be used in English.

Disse-me que viesse ver a casa
He told me to come to see the house.

Diz-me na carta que lhe de este livro.
He tells me in the letter to give you this book.

Quero que ela vá.
I want her to go.

Tenho que ir à escola hoje. (could be counted as both!)
I have to go to school today.

There are also cases in which *que clauses* in Portuguese don't match either English *that clauses* or *infinitives*.

Faz três anos que moro nesta casa.
I have been living in this house for three years.

Similarly, there are cases in which *infinitives* in Portuguese do not match *infinitives* in English.

Ele acaba de ler a sua lição.
He has just read his lesson.

Torno a estudar a lição
I study the lesson again.

Gosto muito de ler.
I like reading very much.

Ele está a estudar a lição.
He is studying the lesson.

Hei de estar ali às quatro da tarde.
I'll be there at four P.M.

PREPOSITIONS

Portuguese makes use of a number of phrasings and constructions which require *prepositions* for apparently arbitrary syntactic reasons, much more so than English seems to do this.

Sempre entram na aula cedo.
They always enter the classroom early.

Aprendemos a falar português.
We learn to speak Portuguese.

Precisa de dinheiro.
She needs money.

Depois de comer.
After eating.

Ela acabou de ler a sua lição.
She has just read her lesson.

Gosto do seu chapéu.
I like your hat.

Vao a falar português.
They are going to speak Portuguese.

Pagamos ao garçon.
We paid the waiter.

O seu livro é igual ao de João.
Your book is like John's.

Um milhão de dólares.
One million dollars.

Hei de estar ali às quatro da tarde.
I'll be there at four P.M.

Um relógio de ouro.
A gold watch.

APPENDIX C
FACTOR SCORES OF THE FOUR GROUPS OF WRITERS

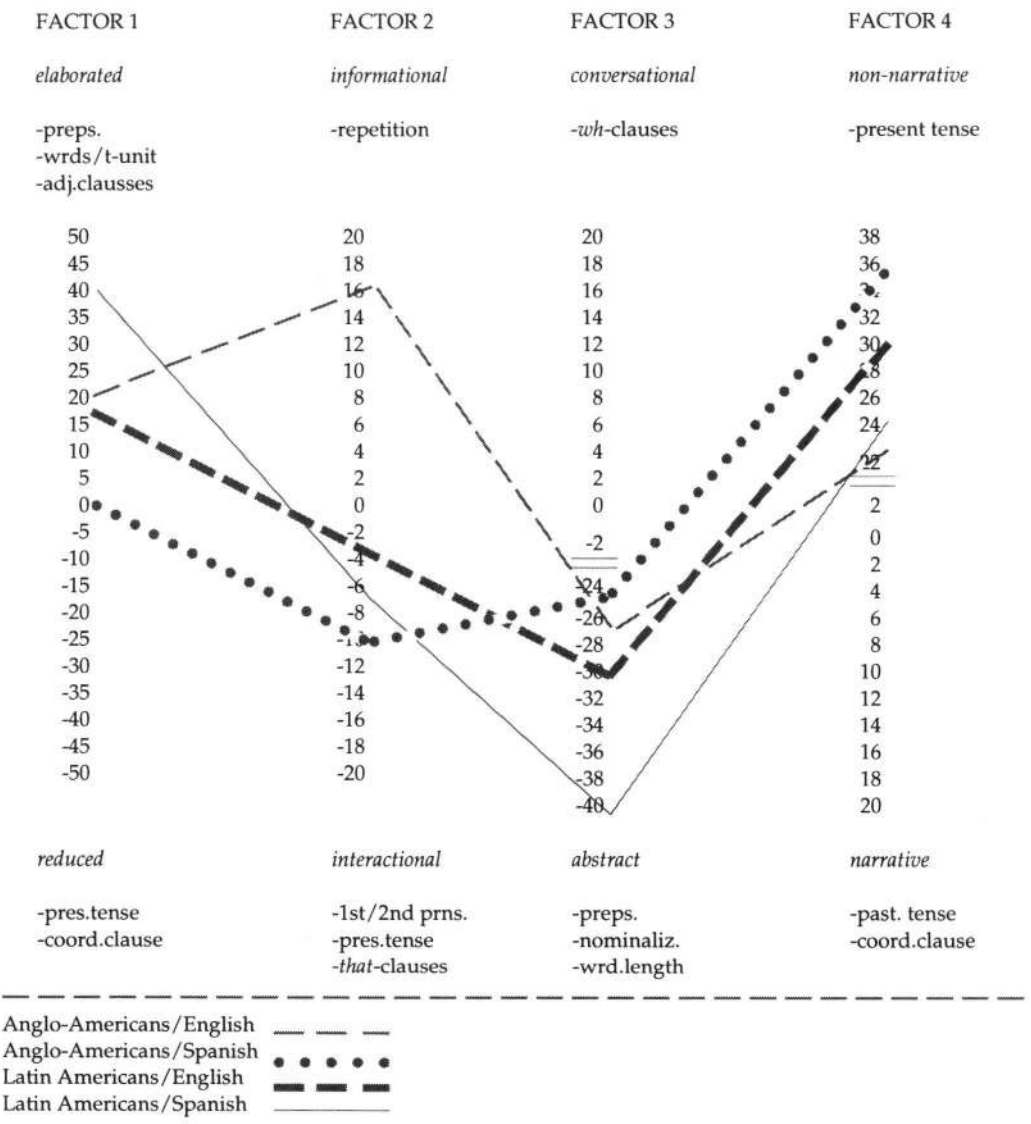

Figure which compares factor means for four groups of writers on each factor and across factors (Lux 1991).