REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
31.Jul.03

3. REPORT TYPE AND DATES COVERED
DISSERTATION

4. TITLE AND SUBTITLE
"VARIATIONS IN WRITTEN ENGLISH"

5. FUNDING NUMBERS

6. AUTHOR(S)
MAJ COLLINS JEFFREY A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CARNEGIE MELLON UNIVERSITY

8. PERFORMING ORGANIZATION REPORT NUMBER
CI02-1189

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

20030822 182

14. SUBJECT TERMS

15. NUMBER OF PAGES
268

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government

Variations in Written English

Characterizing Authors' Rhetorical Language Choices
Across Corpora of Published Texts
July 14, 2003

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Jeff Collins
Carnegie Mellon University

© Copyright 2003 by Jeff Collins. All rights reserved.

Abstract

This book reports on multivariate analyses of two one-million-word corpora of published American English (Francis & Kucera, 1979; Hundt et al., 1999; Hofland et al., 1999).

The study found statistically significant relationships between micro-rhetorical priming features (Kaufer et al., In Press) and high-level rhetorical plans. The study found five groups of co-occurring priming features were affected by text genre at the $\alpha = 0.001$ or better level of significance. These five groups were confirmed as significant across both corpora. These groups were interpreted as language dimensions based upon a qualitative analysis of their usage in the corpora.

The implication is that these five dimensions mark fundamental rhetorical "cut points" in written English, functioning as a heretofore hidden meso-layer linking micro-level linguistic decisions and writers' macro-level rhetorical plans and tasks. The five language dimensions are listed below. The evidence suggests writers of the corpora controlled for these language dimensions in effecting their textual designs.

1. Writing for the Eyes vs. Informing
2. Writing for the Intellect
3. Retrospecting vs. Notifying
4. Instructing
5. Referencing Positive Relationships

This book outlines the background and methods used to define these functional language dimensions and discusses several of the potential implications and research possibilities afforded by these findings.

List of Figures

1.1 Boxplots of scores from 1940s tobacco ads 7
1.2 Tobacco ads from different decades of *Life* magazine 10

2.1 D.R. Russell's Activity Triangle 26

3.1 Scree plot of variances accounted for by unrotated factors 64
3.2 Tukey all-pairwise comparison results for factor 1 77
3.3 Tukey all-pairwise comparison results for factor 2 78
3.4 Tukey all-pairwise comparison results for factor 3 78
3.5 Tukey all-pairwise comparison results for factor 4 79
3.6 Tukey all-pairwise comparison results for factor 5 79

4.1 Eighteen micro-rhetorical priming categories 91
4.2 Heuristic for categories of *thought* cluster 97
4.3 Heuristic for categories of *relations* cluster 107
4.4 Heuristic for categories of *description* cluster 118

5.1 Boxplots showing dimension 1 scores by genre 134
5.2 Boxplots showing dimension 2 scores by genre 150
5.3 Boxplots showing dimension 3 scores by genre 165
5.4 Boxplots showing dimension 4 scores by genre 179
5.5 Boxplots showing dimension 5 scores by genre 189

6.1 Instrumental organization of rhetorical priming categories . . . 202
6.2 Comparison of "editorial" and "reportage" genres 212
6.3 Domain model interaction for "expert" support system 214
6.4 Example web-based interaction 215

A.1 Single Text Visualizer (STV): Dimensional View 225
A.2 Single Text Visualizer (STV): Text View 226
A.3 Multi-Text Visualizer (MTV): Range View 227

A.4 Multi-Text Visualizer (MTV): Map View 228

Contents

Acknowledgements **ii**

Abstract **iii**

List of Figures **iv**

1 Introduction **1**
1.1 Purpose of the study 2
1.2 Significance of the study 4

2 Literature Review **13**
2.1 Multidimensional studies for language analysis 13
2.2 Defining meso-level rhetorical strategies in texts 16
2.3 Using genre as indication of a text's rhetorical plan 23

3 Research Methods **30**
3.1 Overview of the methodology 30
3.2 Text genres of the corpora 56
3.3 Statistical methods 62

4 Rhetorical Priming Theory **80**
4.1 Achieving a goal through different rhetorical plans 80
4.2 Enacting plans through micro-rhetorical decisions 82
4.3 Finding priming effects in text 86
4.4 A hierarchy of micro-rhetorical priming effects 89
4.5 Description of the priming effects 94

5 The Dimensions of Language Variation **127**
5.1 Writing for the Eyes vs. Informing 132
5.2 Writing for the Intellect 149

5.3 Retrospecting vs. Notifying 163
5.4 Instructing 177
5.5 Referencing Positive Relationships 187
5.6 A meso-level rhetorical theory 198

6 Implications and Future Directions **200**
6.1 Instrumental organization of priming categories 201
6.2 Genre analysis and text categorization 206
6.3 Altered view of rhetoric and composition 213

Appendices **217**

A Description of DocuScope **218**
A.1 Frequently Asked Questions 218
A.2 Technical Information 220

B Corpora Changes **229**
B.1 Removing markup tags from the corpora 229
B.2 Separating corpora into constituent text files 231

C Variable Scores by Genre **236**

D Factor Scores by Genre **254**

E Language Dimensions by Genre **257**

Bibliography **263**

1 Introduction

This research study is designed to help us better understand processes of rhetorical invention. Invention has been understood as part of rhetoric since the ancient Greeks and Romans. In the past couple of decades, researchers have made systematic explorations of the cognitive rhetorical processes involved in writing and reading, including invention (Bereiter & Scardamalia, 1987; Hayes, 1996; Haas & Flower, 1988; Kintsch, 1989; Aitchison, 1994). Academics in other fields such as human-computer interaction, communication design, management, and human factors engineering have lately taken up research concerns closely related to long-held rhetorical understandings of invention (Norman, 1993; Johnson, 1998; Schriver, 1997; Geisler et al., 2001). One commonality among these fields—including rhetoric—is that practitioners in the fields engage in what may be considered *productive* design arts: they rely on inquiry and inferences about particular situations to guide decision-making about maneuvers to enact in moving toward finished, acceptable solutions (Schön, 1983; Kaufer & Butler, 1996; Haller, 2000).

1.1 Purpose of the study

A promising theory, first known as "representational composition" (Kaufer & Butler, 2000) and subsequently called "rhetorical priming" (Kaufer et al., In Press), has been offered as a potential description of interchange taking place between writers and readers through texts, the artifacts of rhetorical design processes. This theory, described in chapter 4, is a linguistic and micro-rhetorical framework combining aspects of the theories of I.A. Richards (1991b) and M.A.K. Halliday (1994). This study researches the potential of this new theory to describe rhetorical strategies in written texts.

Kaufer and Butler (1996) developed this theoretical framework based on their observations of how people react to specific language gestures they read in texts and hear in speeches. Kaufer and Butler (2000) proposed this theory as the basis for their effects-based writing pedagogy. Effects-based pedagogies hold potential advantages over the current-traditional pedagogies common in American colleges (Berlin, 1987; Halloran, 1990) because they offer the possibility of more meaningfully focussing writer attention on the design of linkages between text and reader. By making this design task more explicit and available to students, effects-based pedagogies may help motivate consideration of the "perlocutionary effects" of writing (Austin, 1975) and an understanding of language as inducer of "incipient action" (Coe, 1990), filling in what Kaufer and Butler (and others, in various terms, e.g. Noden, 1999; Olson, 1994; Elbow, 1994) argue is an underappreciated, productive communication skill.

There are, however, significant intellectual hurdles still to be worked out: one of the highest is appreciating the links between rhetorical priming theory and rhetorical situations. The basic units of Kaufer et al.'s language theory are the limited variety of patterned, interactive experiences language makes available to readers (listed in figure 4.1, p. 91). These units of language are hypothesized to prime basic rhetorical effects for readers. This study provides an enhancement to this language theory by reviewing empirical evidence of consistent usage of the basic units of the language theory in some common genres of professional writing.

This study develops this bridge by undertaking a multi-dimensional study of two text corpora. This relatively new set of research methods has been applied in several areas of social scientific and language inquiry (Roberts, 1997; Conrad & Biber, 2001; Bybee & Hopper, 2001). At the heart of these research methods lies factor analysis, a multivariate statistical process. Factor analysis yields series of mathematical formulæ accounting for some portion of the variance in multivariate data (Kachigan, 1991), as may be derived from text collections using the methods of discourse analysis (Biber, 1988; Conrad, 1999; Johnstone, 2002). This study set out to find whether statistically significant factors would emerge that suggested links between high-level rhetorical plans and rhetorical priming theory.

1.2 Significance of the study

This study provides our field with a new "meso-level" theory (Neuman, 2000, p. 50) of genre: one that describes genre according to the linkages between macro-level rhetorical strategies and micro-level linguistic choices. Such a development suggests theoretical and practical understandings of how macro-level rhetorical concerns alter the rhetorical invention occurring during the writing process. Such a set of meso-level rhetorical strategies found in written English provides a significant enhancement to the seminal findings of Douglas Biber (1988; 1989) who described variations among written and oral discourse by examining co-occurrences of various linguistic features.

The advances in rhetorical theory provided by this study should prove useful for several applications. Most notably, I believe these factors and the processes used to develop them may guide development of a "domain model" (Wenger, 1987) that will hold potential for enabling cognitive writing tutors (Koedinger & Anderson, 1993; Foertsch, 1995; Steuck et al., 1999) and might serve as a framework for a support system underlying on-line writing labs, an increasingly popular but resource-intensive student support function at colleges and distance-learning centers (Gardner et al., 2001).

In addition, the set of mathematical formulæ describing this study's meso-level strategies (see chapter 3) has potential immediate uses in computer applications ranging from internet search engines and intelligent agents (Rauber & Muller-Kogler, 2001; Roussinov et al., 2001) to new

tools for discourse analysis and corpus studies. It may prove useful, for example, to longitudinal studies being undertaken by researchers at several universities who are interested in writing across the curriculum and long-term portfolio development (e.g. Lunsford, 2002; Benninghoff et al., 2002).

Example analysis to introduce the study

This section provides a short example to help you see the broader questions being answered by this study. This example shows how the language theory underlying the study may be useful for providing insight on rhetorical plans, represented in the study by text genres.

Kaufer et al.'s (In Press) theory of rhetorical priming (described in chapter 4) provides a theoretical framework for understanding everyday language. It's based on the notion that small, systematic differences in rhetorical language choice may accumulate into large, important differences. Kaufer et al. developed a coding scheme and computer software (described in appendix A) based upon this language theory that can be used, along with multivariate statistics, to examine the usages of rhetorical priming in collections of texts.

There are 18 categories of rhetorical priming described by the theory. For simplicity this example considers only six of them, shown in the left column of table 1.1. Using the software, I counted the percentage of text devoted to each of these priming categories in a collection of tobacco advertisements from the 1940s (see Wooden et al., 1998, for the collection).

TABLE 1.1: Comparison of sample tobacco ad Copy from *Life* Magazine

	1940s ($n = 28$)		1950s ($n = 23$)		ANOVA results	
Priming Category	mean	SD	mean	SD	F	p
First Person	0.436	1.030	0.410	0.846	0.010	0.924
Inner Thinking	3.409	1.623	3.403	2.243	< 0.001	0.991
Think Positive	3.551	1.367	5.477	3.151	8.540	0.005
Think Negative	0.980	1.165	0.487	0.754	3.060	0.086
Linear Guidance	4.636	2.031	4.596	1.387	0.010	0.936
Word Picture	7.533	2.350	7.911	2.660	0.290	0.593

$df = (1, 49)$

The mean and standard deviation scores for each priming category from this collection are presented in the second and third columns of the table. The mean score indicates the average percentage of text devoted to that priming category and the standard deviation indicates the variability among the texts on that category.

Another way of representing these data is by using a boxplot to show the texts' scores on each category, as I've done in figure 1.2. The figure shows the scores of all the texts on each priming category. The median score is indicated by the horizontal line crossing the box and the variability is indicated by the overall length of each plot. As you can clearly see from the figure, some of the priming categories have a lot of variability (long boxplots) while others have less variability (short boxplots).

For example, we see most of the collected tobacco advertising used very little "first person"—the median score is low (0.0%) and the boxplot is quite short. At the same time, the advertisements used quite a bit more

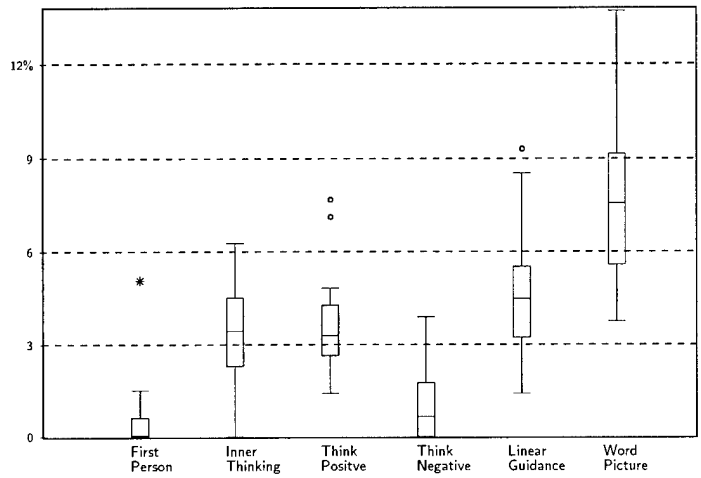


FIGURE 1.1: Boxplots comparing scores on six priming categories from tobacco ads of 1940s *Life* magazine

"think positive" priming—its median is higher (3.3%)—yet the boxplot is also relatively short, indicating the texts use a fairly consistent amount of this priming category. On the other hand, there is much more variety of " word picture" priming in the texts: some of the writers used a lot and others a little.

What does this information tell us? It may indicate the importance of "first person" and "think positive" priming to this genre of writing: the writers of the collected ads seemed to control for these of priming categories while letting some of the others roam more freely. Not *all* of the ad writers controlled these priming categories—there are outlying usages of "first person" and "think positive," indicated by the asterisk and circles above the top whisker of the boxplot—but *in general* it appears the writers of these collected ads used a consistent amount of priming in some of

the categories.

The idea behind the study presented in this book is that there may be systematic patterns to these differences that are based upon a texts' genre. To see whether this is the case, the study uses methods of corpus linguistics to compare collections of texts to one another to find out whether there are differences that can be detected (there are, as you will see in chapter 3) and whether these differences are systematic and might be qualitatively meaningful (they are, as you will see in chapters 5 and 6).

But there's an important experimental consideration that must be acknowledged whenever one begins to use genres to represent static categories of texts: Rhetoricians interested in the study of genre have consistently indicated that many aspects of the rhetorical situation affect genre. Genres are not static categories of text, but are themselves built of texts generated by authors who are responding to previous texts and to their current writing situations (Bazerman, 1994b; Swales, 1990; Berkenkotter & Huckin, 1995; Freedman & Medway, 1994). These studies have demonstrated that genres guide writers—writers must operate within and against a genre when composing texts—but genres are also built by the writers acting in the situated moment and evolve (or dissipate) with each new writing.

To see this in action, let's take a brief look at another group of texts, this time tobacco advertisements of the 1950s. The summary statistics for these texts are provided in the middle columns of table 1.1. We can compare these texts to the texts of the 1940s quantitatively using a statistical procedure called "analysis of variance" (ANOVA). This procedure

is useful because it compares the scores of the groups, takes account of each group's variability, and then provides an indication of how confident we can be that the differences between the scores is due to group membership and is not the result of random variation. The ANOVA process yields a statistic known as the F-score and another known as the p-value. These statistics describe the confidence warranted by the results of the comparison. A low p-value indicates a low probability the difference is due to random error.

By looking at the results of this test (the right-hand columns of table 1.1), we can see most of the priming categories of the tobacco advertising are not significantly different between the two decades (the p-values are quite high). This is an interesting result because it suggests the authors of tobacco ads used most of the priming categories consistently across the decades.

But also of interest are the changes that occurred, as indicated by the low p-values for the "think positive" and "think negative" priming categories. These values indicate the writers' usage of these two categories changed between the decades the texts were written.

What is going on here? Well, the statistics can only tell part of the story. To really see what is happening and to begin to explain these statistical results, we have to dig into the texts. This is where qualitative analyses of the collected texts become vital (as you will find in chapter 5).

We can see from table 1.1 that the tobacco ad copy of the 1950s uses, on average, more strings priming positive feelings than the advertisements of the 1940s. We experience this same result when reading the texts. We

(a) 1940s (b) 1950s

FIGURE 1.2: Tobacco advertisements from different decades of *Life* magazine

find, for example, the following 1950s text accompanying figure 1.2(b), a picture of a male lifeguard holding a cigar:

> He alone rules the happy chaos of a million beach-goers every summer. It's a job that takes skill and patience. And like men everywhere, he makes the going easier by enjoying the cigar that's one in a million—mild, yes tastefully mild—Dutch Masters.

Contrast that image of male responsibility over "happy chaos" with the

image of responsibility conveyed by the following 1940s ad copy, which accompanied the image shown in figure 1.2(a):

> He's a Bombardier. He's the business man of this B–17E bomber crew. His office is the "greenhouse" of transparent plastic in the nose of the ship. And he works there on split-second time. But when those office hours are over—well, just look below and watch him enjoying a Camel—the favorite cigarette on land, sea, and in the air.

Similar differences are found throughout the collected texts. The 1940s ads use, on average, twice as much "think negative" priming and a third less "think positive" priming than the 1950s ads do. These differences come through in the verbal images of wartime duties as well as in ad copy that attempts to claim new features or manufacturing processes that overcome some of smoking's deleterious effects—effects that for some reason are rarely mentioned in the 1950s advertising.

This short example demonstrates how carefully examining the rhetorical priming in texts provides interesting possibilities for finding similarities as well as differences among text genres, as rhetorical scholars interested in genre would surely predict.

But the question of this study is whether or not there is sufficient stability among text genres to provide us with useful insights on rhetorical language usages, insights that would help us see how authors bridge the divide between high-level rhetorical goals and the micro-level linguistic decisions that would be needed to effect those goals in text. The study

presented in this book uses quantitative and qualitative means to offer an affirmative answer to this question.

2 Literature Review

This chapter reviews the relevant literature supporting the current study. It covers previously published research supporting understanding in three areas of scholarship:

- the utility and acceptability of corpus-based multidimensional methods of language analysis;
- the interest in understanding meso-level rhetorical strategies;
- the feasibility of using the corpus-based research methods to define such rhetorical strategies.

2.1 Multidimensional studies for language analysis

With the widespread availability of fast computers, interest has grown in using language corpora as an aid to understanding language usage (Bybee & Hopper, 2001; Biber et al., 1998; Conrad & Biber, 2001; Nickerson, 2000; Palmquist et al., 1997). Susan Conrad (1999) demonstrates the characteristics of corpus-based research and illustrates how such studies

help meet the needs of writing students and teachers by allowing comparisons of language characteristics across varieties of texts. She writes, "Studies that include statistical analyses and frequency data can identify strong patterns in language use that we do not recognize intuitively" (p. 2).

Joan Bybee and Paul Hopper (2001) make a related argument in the introduction to their recent collection of linguistic studies. They argue detailed (though implicit) tracking of frequencies and probabilities of exposure may be "an important factor" in how people learn and use language. The essays in their collected volume demonstrate the important implications of frequency studies for language researchers.

This is not to suggest such studies are a panacea for writing researchers. Charles Perfetti (1998), for example, cites the drawbacks of theoretical research that exhibits an overreliance on frequency of co-occurrence. Sven Birkerts (1994) bemoans the possibility of uncritical acceptance and reliance on computerized mechanisms for accessing and analyzing texts, warning of the potential for negatively altering our cognitive relationship with the written word (cf. Partington, 1998; Bracewell, 1999; Herrington & Moran, 2001). Nonetheless, the literature suggests carefully conducted frequency studies do enable powerful insights that would otherwise remain inaccessible, as is amply demonstrated by a number of recent books and articles making use of such methods (e.g. Downs, 2002; Vande Kopple, 2002, 1998; Honeycutt, 2001; Hyland, 2001; Conrad & Biber, 2001; Haswell, 2000).

Several of these recent investigations use corpus-based methods to

link macro- and micro-level language concerns in ways similar to the current study. The work of Ken Hyland (2001), for example, lends support to the substantial literature showing how writers use various linguistic features to shape their texts in anticipation of specific reader expectations. He uses the methods of corpus analysis to examine selected texts from research articles in ten journals across eight different disciplines, focussing his study on explicit features of reader orientation to gain insight on rhetorical engagement in academic discourse. Similarly, William Vande Kopple (1998; 2002) employs these research methods to follow up on Charles Bazerman's (1994a; 1994b) seminal works on genre systems in patent documents and scientific discourse. Vande Kopple's (2002) analyses buttress Bazerman's findings, but also highlight alternative possibilities consistent with concepts from M.A.K. Halliday's (1994) functional grammar. Likewise, Richard Haswell (2000) shows the benefit of quantitative frequency studies—including the use of multivariate factor analysis—for longitudinal rhetorical investigations, finding suggestive, otherwise hidden relationships among the assessed variables in the portfolios of writing students as they progressed through their college careers.

Such published studies demonstrate these research methods' potential value, as well as their acceptance by the field. Several characteristics of such corpus-based analyses account for their reliability, as Biber et al. (1998, p. 3–5) suggest:

- they are empirical, analyzing actual language patterns found in a

"principled collection" of texts;

- they are consistent, making use of computers and statistical techniques to assist the researcher in finding significant patterns;
- they are humanistic, capitalizing on both qualitative and quantitative analytical techniques.

Employing these methods for this study enables similarly powerful insights into the particular rhetorical behavior that is this investigation's focus, as I will show in the following sections and chapters.

2.2 Defining meso-level rhetorical strategies in texts

This section presents some of the literature considering the stochastic nature of composition and how authors might accommodate such uncertainty through linguistic conventions. Along with David Kaufer (Kaufer et al., In Press), I have been considering whether it may be the stored units of language usage that instantiate larger rhetorical goals during "languaging" (Becker, 1991). Perhaps people learn micro-level rhetorical effects through experience and then use these micro-level effects as substrate for building larger effects. That is, perhaps people create macro-level rhetorical effects by keeping implicit mental account of the intensity, frequency, and blending of their substrate-usage in particular pieces of discourse. Bybee and Hopper (2001; and also Biber et al. (1998, pp. 135–171)) suggest similar possibilities in regard to grammatical considera-

tions: "Linguistic material cannot accrue frequency effects unless the brain is keeping track of frequency in some way..." (Bybee & Hopper, 2001, p. 10).

The current study is motivated by the idea there will be fruit in pursuing an understanding of how writers accommodate the stochastic nature of rhetoric and that Kaufer et al.'s (In Press) "rhetorical priming theory" provides a potentially useful description of the micro-level units—or substrate—of rhetorical language behavior. This study is designed to test this potential by attempting to find patterns of the theorized rhetorical substrate that relate to the high-level goals found in actual samples of professional writing. This idea of linking substrate usage to broader rhetorical concerns is found in several of the studies I outlined in the previous section, and is also partially articulated by Richard Coe (1990; 1994b) and Todd Oakley (1999) in their considerations of rhetoric and the human rhetorical potential[1] (cf. Nystrand, 1986). Finding such patterns of substrate usage would demonstrate the potential of Kaufer et al.'s theory for defining useful units of rhetorical analysis and of pedagogical interest. At the very least, this study provides our field with empirical in-

[1]Oakley's paper goes beyond characterizing a "rhetorical potential," toward attempting to define a full-blown "cognitive rhetoric," a project that has been set out by Mark Turner and other cognitive researchers in volumes with titles such as *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities* (Fauconnier & Turner, 2002). While I acknowledge the interest and potential value in attempting such connections, this study does not attempt to extend beyond improving our understanding of writing, leaving unspecified any generalized connections between writing, language and thinking. I address the reasons for this purposeful limitation later in this book.

sight on rhetorical invention, one aspect of the rhetorical design process.

There are intriguing parallels between rhetorical processes and the processes used by design practitioners (see Schön, 1983; Kaufer & Butler, 1996). Cynthia Haller (2000) has recently articulated some of these parallels between writing and design: "in both, writers/designers discover solutions to problems, conscious self-monitoring tends to produce greater success, processes are recursive and cyclic, and attention to the needs of readers/users is necessary" (p. 355). According to Haller, these parallels have led some rhetoricians to claim an apparent role for rhetoric in the processes of design and, also, a place for some of the well-understood design processes in rhetorical theory. But however suggestive the parallels may be between the activities of writers and designers, Haller points out that such claims have been theoretically rather than empirically based and that "we still know little about how processes of rhetorical invention are played out in particular design contexts" (p. 356).

Although the activity of rhetorical invention has been sometimes considered to be limited to the early stages of writing (e.g. Steuck & Rowley, 2000), Cicero (1954) implies invention is employed throughout the creative process, aimed always at shaping the discourse to achieve the desired end (I.ii.2–3, I.iii.4; cf. Sloane (1989); May & Wisse (2001, pp. 29–32)). Likewise, even in the earliest models of their composition process, Hayes and Flower (1994) stipulate non-sequential, *overlapping* composition activities, addressing the issue specifically with this caveat:

> We should caution the reader not to interpret our model as a

> stage model. We are not saying that writing proceeds in order through successive stages... The model is recursive and allows for a complex intermixing of stages. (p. 178)

Hayes and Flower and others (e.g. Aitchison, 1994; Bereiter & Scardamalia, 1987) show that strong separation of writing into stages of activity does not conform to observed writing behavior. There are many reasons for the overlapping of composition activities, but the one of interest in this study is the stochastic nature of writing, a reason suggested by Coe (1994b) in a rather speculative footnote to an essay considering the nature of writing and writing genres: "The writing process is always purposive, hence not random, but stochastic (from the Greek, to aim, to guess) because it is more often goal-seeking or goal-oriented than goal-directed..." (p. 167n5). This stochastic nature is a particular problem for writers because texts are so clearly subject to negotiated meaning. As many researchers have shown (Haas & Flower, 1988; Tomlinson, 1990; Gage, 1991; Olson, 1994; Flower, 1994), readers bring a host of experiences and attitudes to texts that help shape their readings. For example, N.L. Gage (1991) found that readers consistently judged social and educational research findings as "obvious," which tended to work against writers' attempts to convince readers that a breakthrough had been found or that further research should be funded. Even in writing situations where readers are well-known to writers, the writers must make educated guesses about how the reader will take the text and what the writer should do about this uncertainty while composing.

Rhetoricians have long described and attempted to provide guidance for dealing with this stochastic nature during linguistic performances. Coe (1990, p. 49)—following Richard Burke—suggests authors struggle to induce "incipient actions," a term used to describe the "shared attitudes" or "leanings" that bring readers toward cooperative action, not just at the end of the text, but also during the reading of text. This is similar to Perelman and Olbrechts-Tyteca's (1969, pp. 65–76) concepts of "sharing" that they argue are necessary for the "unfolding of argument" in *The New Rhetoric.* This "sharing" takes place (or does not take place) regardless of whether the audience is ultimately persuaded by the text. For example, an argument to ban a chemical to save birds requires more than just a presentation of evidence showing the chemical harms the birds: the argument also requires the writer to induce shared understandings of what constitutes valuable evidence and acceptable premises, such as birds holding value for the world. Authors induce such incipient actions through the linguistic performances they instantiate in text during the composition process (cf. Freadman, 2002).

The ability to do so requires aspects of what Todd Oakley (1999) has defined as the "human rhetorical potential." Oakley's argument for such a potential is complex and is seated in biological and cognitive capacities that allow and, arguably, necessitate symbolic behavior among humans. While Oakley acknowledges that such neurological considerations are generally removed from most rhetorical studies, he insists these considerations "complement" long-held rhetorical assumptions (see also Foertsch, 1995). For example, Oakley uses a definition of memory from

Gerald Edelman's 1992 *Bright Air, Brilliant Fire,* describing memory as "the ability to repeat a performance" and suggesting that routinized linguistic forms are "a computationally cheap way to hold thoughts steady in working memory" (Oakley, 1999, pp. 109–110). Oakley suggests such neurological capacities provide humans with the potential for rhetorical thinking.

He finds suggestions of such capacities in different rhetorical theories, noting, especially, how they are engaged by Perelman and Olbrechts-Tyteca's concept of "presence" and Burke's concept of "identification." Oakley argues that both of these concepts and the rhetorical theories of which they are part "focus on the study of *constructions,* form-meaning pairings arising from the matching of a conventionalized schema with particular elements in a particular scene" (p. 103; see also Enos, 1990; Hanson, 1997). Oakley argues we rely on such a "conventionalized set of constructions" for rhetorical behavior. That is, we rely on shared linguistic conventions in order to prompt others to share our mental simulations. As evidence, Oakley cites articles by cognitive researcher David Kirsch in which people are observed manipulating items in their surroundings to help them and their collaborators improve memory and accomplish tasks (cf. Norman, 1993). Oakley infers from these observations that people attempt to produce a "cognitive congeniality in ourselves and others for the purpose of affecting thought and action" (p. 100). This is a notion similar to Coe's (1990) "incipient actions."

Theresa Enos (1990) describes a similar rhetorical interchange as she offers her conception of "consubstantiation." Her conception is apropos to

this discussion insofar as it focusses on language as a mediating means for achieving identification. Her essay is specifically concerned with arguing for the existence of a generative *ethos* that emerges from text as a result of the text-audience and text-rhetor relationships (a central tenant of James Berlin's (1987) "epistemic rhetoric"). In arguing for such an emergence, Enos considers the balance between ancient conceptions of the rhetor as the master of audience and some contemporary post-structuralist conceptions in which the reader brings all meaning to text. Enos argues that the latter theories have "led us too far away from the writer's dialogic imagination that invites the reader... to create meaning and thus become part of textual interanimation" (p. 103; see also Flower, 1988, 1994). Enos provides evidence for what she calls the "audience/reader pair," requiring an interanimation by which writers create a version of an audience in text and readers identify this evoked audience and become (or refuse to become) just such an audience for the text (p. 104, cf. Ong, 1975). Using Burke's term "identification" to describe the outcome of such processes, Enos introduces the term "consubstantiation" (also from Burke), defining it so as to emphasize the necessity of dialogic action taking place as the writer discovers possibilities for identification as he or she becomes an audience for the text being produced. Enos argues that "In this way, among others, the writer has created identification with the text, an identification that is mediated through language and that would be easily accessible by a reader" (p. 110).

This is consistent with Oakley's (1999) argument about aspects of an inherent rhetorical potential and our ability to "select and attend to par-

ticular phenomena in the world while ignoring other phenomena... [and to] develop routines for representing those choices to others that allow [them] to 'read' it reliably (although perhaps not perfectly)" (p. 102). Oakley demonstrates how linguistic conventions manifest the enriching and limiting of rhetorical invention,[2] driving selective attention to the management of particular constructions and causing the amalgamation of ideas into shared, identifiable language patterns. Authors must control these patterns to induce publicly-traded projections and it is these patterns—patterns that might be considered to comprise a rhetorical substrate—that Kaufer et al.'s theory of rhetorical priming may describe. This study aims to assess whether there is a consistent usage of these particular language patterns by using methods of corpus linguistics.

2.3 Using genre as indication of a text's rhetorical plan

To explore the possibility of a rhetorical substrate described by Kaufer et al. (In Press), I propose using statistical methods to seek a relationship between micro-level linguistic features in texts and text genre, which this study takes as an indication of the author's macro-level rhetorical plan.

The term "genre" is slippery because it is used in different fields to indicate different categorizations of texts, as indicated in Swales' (1990) overview of the subject. For example, some scholars might divide litera-

[2]Oakley does not use "invention," but, instead, uses the word "thought." I've opted to narrow the claim here to avoid the charge of accepting the Sapir-Worf hypothesis.

ture into short stories, novels, and poetry, calling each a "genre." Or they might categorize poems into "genres" based upon text features such as topic or rhythm or rhyme scheme or some combination of text features (e.g. sonnets versus limericks). Still others might categorize texts into genres based upon rhetorical forms such as the eulogy, the apology, the letter of recommendation, or the business progress report.

In this study, I chose to accept the genres of texts in the Brown corpus (Francis & Kucera, 1979) as an indication of the writers' rhetorical plans. In doing so, I follow a notion about genre first articulated by Carolyn Miller (1984) and taken up by other rhetorical scholars (Bazerman, 1994a,b; Swales, 1990; Paltridge, 1997; Berkenkotter & Huckin, 1995; Fahnestock, 1993; Cooper, 1999; Coe et al., 2002). Miller defines genres as responses to perceived, recurrent rhetorical situations and views genre as "a point of connection between intention and effect, an aspect of social action" (Miller, 1984, p. 153). Although this approach to genre has been the basis of much scholarship, how the approach should be applied to the study of texts is a contested question (Freedman & Medway, 1994; Geisler, 2001) and has been implicated in worsening what Richard Coe (1994b, p. 161–163; 1994a, p. 183–185) describes as a "false dichotomy" between focussing attention on writing forms or on writing processes. Coe explains how this form/process dichotomy is ultimately artificial and damaging to our field. He suggests we should, instead, understand the structures of writing genre as corresponding to the nature of the rhetorical situations in which writers find themselves in the process of writing. Citing several of his contemporary scholars and building his argument from

Richard Burke's (1966) *Language as Symbolic Action* and Karen Burke LeFevre's (1987) *Invention as a Social Act,* Coe suggests scholars should consider genres as "fossilized rhetorical processes," available for analysis, but always with the understanding that genres are both "constraining and generative" to writers and, therefore, common and useful to scholars with interests on both sides of the supposed form/process dichotomy (p. 185).

A parallel explanation of this perceived dichotomy is more recently articulated by Cheryl Geisler (2001) who proposes that the many rhetorical scholars who insist on a social-action theory of genres (e.g. Miller, 1994; Freedman & Medway, 1994) have resisted focusing on text forms because they "fear" such a focus might "reify the text as object, normalizing and stabilizing its effect" (p. 297). Geisler argues this fear arises from "an outdated view of objects, one that takes them as divorced from or predating social use" (p. 298).

She counters this view by drawing on activity theory (see, e.g., Hutchins, 1996; Bazerman & Russell, 2003) to demonstrate that providing a focus on text forms enables more complete rhetorical scholarship because it allows scholars to begin to account for "recurring formal features that often identify specific texts as exemplars of one genre or another, a fact that most social-action theories mention but characterize as accidental" (p. 297; cf. Charney & Carlson, 1995; Cooper, 1999). Geisler borrows D.R. Russell's (1995) formulation of the activity triangle (figure 2.1) to help her explain the mutually-supporting stances on text she recommends for scholarly studies of genre.

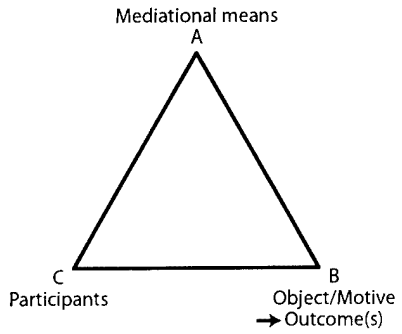


FIGURE 2.1: D.R. Russell's Activity Triangle (adapted from Geisler (2001), p. 298).

Scholars taking genre as a situated activity (following Miller, 1984) have often placed texts at (A) in the diagram, taking texts simply as mediational means through which the participants (C) achieve their outcomes (B). In other words, the situated activity is completed through the use of text, but the text is not the objective of the activity. This stance is found, for example, in Dorothy Winsor's (2000) study of the use of texts by blue collar technicians and white collar engineers at a farm equipment manufacturer. In studies like Winsor's, altering the research stance to make text features the primary object of study would not be appropriate because the political nature of the situation would not be fully available from an account of textual features.[3]

However, as Geisler (2001) points out, texts are not just mediational

[3]In this distinction and its methodological implications lies some of the justification for scholars aligning themselves with "genre studies" (Freedman & Medway, 1994; Miller, 1994), a field which some argue should emerge from rhetoric and concern itself more with broader societal issues and criticism than with text and other forms of discourse.

means. They are also the outcome of activity, which places them at point (B) in the diagram, at least sometimes. Geisler provides the example of an engineering firm in which the outcome of a meeting is to be a document specifying design requirements. While such a text might itself be viewed as mediational means of a larger activity system, examining the features of the text itself might provide insight into the meeting or even into a group of similar meetings. That is, by studying the features of texts that are the outcome of these engineering design meetings, the researcher gains insight into the particular activity and might have a better accounting of the role of the text as a driving motive of the people engaged in the situated activity.

We see such a research stance taken in studies by Bazerman (1994b), for example, whose focus on the changes in textual features in genres over time provides insight into the mechanisms for ongoing scientific development: his suggestion (p. 99) is that features of the generic form of the text being produced partially directs the behavior of those producing it. This is similar to Coe's (1994b, p. 185) suggestion that genres can be taken as "fossilized rhetorical processes" that are "constraining and generative" for writers and their analyses should afford researchers productive insights.

Using this notion of genre as fossilized rhetorical situation, this study attempts to develop an understanding of the invention–languaging relationship by seeking the presence of consistent differences in the micro-rhetorical priming features among 14 of the genres in the Brown and the Friedburg–Brown corpora (Francis & Kucera, 1979; Hundt et al., 1999).[4]

[4]The genres in the corpora to be considered in this study are presented in Table 3.1,

In statistical terms, this means the null hypothesis for this study is an expectation of finding no significant differences (at the $\alpha = 0.05$ level) among the text genres.

There are well-known empirical limits of such post hoc rhetorical analyses for building a theory of language use (Flower, 1994; Bracewell & Breuleux, 1994; Johnstone, 2002). This study examines the published *results* of the writing process and is, therefore, limited in scope and has nothing concrete to say about any particular "steps" within real-time discourse production. Although this is certainly a limitation to the applicability of my findings and acknowledging such limitations is somewhat unusual for rhetorical theorists using genre, accepting such limitations is in keeping with several other studies interested in describing the outcomes of rhetorical processes (e.g. Van Dijk, 1986; Nickerson, 2000; Vande Kopple, 2002) and is appropriate to the early stages of language theory development (Conrad & Biber, 2001; Perfetti, 1998). The results of this examination is not subject to the unreliabilities of retrospective accounting by discourse participants, nor to any invalidating alterations of the processes that necessarily occur with some methods of direct observation of language behavior (Bracewell, 1999; Faigley et al., 1985; Swanson-Owens & Newell, 1994). Furthermore, limiting this study to an examination of completed texts is beneficial for attempting to characterize mechanisms of interchange between writer and reader than would be a real-time observation of the writing process. By using corpora of professionally-written

page 33. The genre called "Miscellaneous" will not be used for developing the meso-level theory, for reasons described in section 3.1.1 (p. 32) of the chapter on research methods.

texts divided into systems in which the texts make sense (the genres), this study is more likely to achieve a valid understanding of the rhetorical behavior behind the texts' production, as Coe (1994a) suggests:

> . . . a complex stochastic process like writing is not best described by attempting to follow temporally as writers write—because the whole point is that, efficiency aside, a writer may work in *any* order as long as the final product is satisfactory. The process is best understood by describing not a writer's behavior, but the system within which that behavior makes sense. . . (p. 167n5)

While taking texts and genres as fixed points in the study of rhetorical activity is certainly not appropriate for all research studies, I have attempted to demonstrate why it is acceptable for this study.

3 Research Methods

This chapter describes the research procedures and materials used in the study. The methodology closely follows Douglas Biber's (1988) study of linguistic variation in oral and written discourse. In addition to an overview of the methodology, this chapter describes the two, 500-text corpora used in the study, known as the Brown and Friedburg–Brown corpora of written English (Francis & Kucera, 1979; Hundt et al., 1999). This chapter also introduces the rhetorical language theory (Kaufer et al., In Press) underlying the analysis of the corpora.

3.1 Overview of the methodology

An exemplar for corpus-based language research is the work by Douglas Biber (1988) of linguistic variation across oral and written discourse. His study gathered a broad corpus of works and used statistical methods to examine patterns of co-occurrence within the texts. Biber counted "the widest possible range" (p. 72) of linguistic features as the basis of his exploration. Reviewing a bevy of linguistic studies on differences between spoken and written discourse, Biber chose to explore how 67 different

text features were associated with one another in the texts of his corpus. This method enabled Biber to define what he calls "functional dimensions" (pp. 91–92) of linguistic variation that explain the co-occurrence patterns among the text features.

The current study adopts a similar methodology, but toward a somewhat different end. The question this study sets out to answer is whether or not differences in text genres (described in section 3.2) may help explain different usages of specific micro-rhetorical language features (described in chapter 4).

To answer this question effectively, the corpora in this study are limited to written texts that had been separated into different genres by human readers.[1] This enables the application of the hypothesized language theory to a group of texts written for a variety of purposes and classified independently of the theory.

As an organizer for describing the research procedures, this section follows an adaptation of Biber's (1988, p. 64) methodological outline:

- Collect texts and convert them to machine-readable form
- Identify the potentially important linguistic features
- Count the occurrences of the features in the texts
- Perform multivariate factor analysis

[1]Biber used similar written text genres in his 1989 study, but applied the dimensions he had developed in his 1988 study of the mixed corpus of oral and written discourse. Biber's studies were also based on texts of British (rather than American) written English (see Biber, 1988, pp. 65–71).

- Determine whether the factors are affected by genre
- Interpret the factors as textual dimensions

As indicated in chapter 2, the methods of corpus analysis are becoming commonplace and are well described in corpus linguistics textbooks, such as Biber et al. (1998). Several references, in particular, guided this study: Biber's (1988; 1989) works and the collection of multi-dimensional language studies gathered into the volume by Conrad & Biber (2001) were helpful, as were the statistical handbooks by Kachigan (1991) and Stevens (2002); likewise, a chapter by Bryant & Yarnold (1995) was useful in that it provided further detail regarding principal-components and factor analysis from the perspective of the American Psychological Association.

3.1.1 Collect texts: the Brown and Frown corpora

As the primary corpus for this study I used the Brown Corpus of texts (Francis & Kucera, 1979). I used a second corpus, the Freiburg–Brown Corpus (Hundt et al., 1999)—referred to as the "Frown corpus"—to help verify the results of the statistical processes. The use of the second corpus enabled me to gauge the stability and reliability of the qualitative interpretations of the factors, an important standard for judging the potential long-term impact of the study presented here.

The two parallel corpora each consist of 500 texts that were categorized into 15 different genres (listed in table 3.1). Each of the text samples appeared in print in a professional publication during the years the cor-

TABLE 3.1: List of the genres and the number of texts (n) per genre in the corpora

Reference[a]: Text Genre	n
A: Press Reportage	44
B: Press Editorials	27
C: Press Reviews	17
D: Religion	17
E: Skills and Hobbies	36
F: Popular Lore	48
G: Belles Lettres, Biography, Memoirs	75
H: Miscellaneous	30
J: Learned	80
K: General Fiction	29
L: Mystery and Detective Fiction	24
M: Science Fiction	6
N: Adventure and Western	29
P: Romance and Love Story	29
R: Humor	9

[a] The reference letters were assigned by the compilers of the Brown corpus and are used in tables and graphs throughout to refer to the genres.

pora were compiled (1961 for the Brown; 1991–92 for the Frown). Each corpus has approximately 1 million words, with roughly 2,000 words in each of its 500 texts. I made use of widely-available electronic versions of the corpora on CD-ROM (Hofland et al., 1999). Appendix B describes the details of the adjustments needed to prepare the corpora text files for use in the study.

The list of text genres in the corpora was drawn up by a conference of linguistic scholars meeting at Brown University. The texts in the corpora were written to serve a broad range of purposes from academic texts to

newspaper articles to fictional mysteries and westerns.

The text selection procedure used to compile texts for the corpora was designed to attempt to gather a random, representative sampling of all published works, making the corpora ideal for linguistic analyses of a variety of language phenomena (the latest bibliography of such studies includes nearly 300 works making use of the Brown corpus). The Frown corpus was developed at the University of Freiburg, Germany, with the aim of paralleling the makeup of the original Brown corpus as closely as possible, but with contemporary texts. I describe the genres of the corpora in more detail in section 3.2 (p. 56).

There are several advantages afforded by using the Brown Corpus as the primary corpus for this investigation. First, the diverse texts in the corpus were collected and categorized by a group of scholars unfamiliar with the language theory of interest in this study. This achieves an independence between the collection of texts and the theory being tested, an independence that is further enhanced by the randomized text-selection procedure. Although for some applied linguistic research such independence is not necessary (e.g. Conrad, 1999), in this case the independence alleviates the possibility of selection bias and tautological explanations of the study's findings (Keppel et al., 1992; Neuman, 2000).

A second benefit of using the Brown Corpus is that the genres were suggested and the texts categorized by the group of linguistic scholars: the categorization of the texts into the genres indicates the texts within each genre share enough recognizable features that the body of scholars agreed each text makes sense within the bounds of its settled-on genre.

This is important because the categorization is a result of an overall reader-response, and is not based on any linguistic feature in particular.

Third, the corpus is comprised of texts written by professional writers and collected to represent a randomized sample of all published work during the year. Although the texts are not all completed works (each text is limited to approximately 2,000 words), each text is an outcome of a writing process and at least partially satisfied an objective for its writer. In other words, each text is its author's response to a particular rhetorical situation. By using such a corpus of texts divided into systems in which the texts make sense (the genres), the study is likely to achieve a valid description of some aspects of the rhetorical behavior behind the texts' production, as suggested by Coe (1994a) and other rhetorical scholars interested in genre (refer to section 2.3 for further discussion of the rhetorical interest in the study of genre).

Of the 15 genres in the corpus, I used the texts from 14 in seeking to develop the meso-level theory in answer to the motivating question of the study. The additional genre, "miscellaneous," was not used because the texts collected in this category do not meet a rigorous definition of genre, at least not in the rhetorical sense. The "miscellaneous" texts were simply too dissimilar from one another for the scholars who compiled the corpus to consider the group anything but a catch-all category. In other words, texts that did not fit well elsewhere were put into the "miscellaneous" genre. Therefore, the rhetorical goals and the situations surrounding the composition of these texts would be too varied to be well understood and, thus, I excluded this genre from the theory-development

parts of this study.

3.1.2 Identify text features to count: A theory of rhetorical priming

In Biber's (1988, pp. 52-55) study, an extensive literature review of traditional linguistic features was made in order to attempt to consider a wide range of features that might contribute to variation among oral and written discourse. In some senses such an investigation is like a fishing expedition, casting a wide net in hope of capturing interesting co-occurrences that the researcher will be able to interpret meaningfully (see Perfetti, 1998; Bybee & Hopper, 2001, for a description of possible confounds in adopting such a methodology).

The present study is different in that it takes up a fully-formed rhetorical theory and tests whether or not the narrow range of language features hypothesized as part of the theory is related to the actual language usages in the corpus. In statistical terms, this means there's a null hypothesis for this examination: the co-occurrence of the language features in the rhetorical theory may not be related to the a priori text groupings (a possibility evaluated in section 3.3.2).

Before introducing the rhetorical theory and its component language features, I want to reiterate what studies of such co-occurring features attempt to accomplish. Clearly people normally don't speak and write with the aim of using any particular feature: they speak and write to accomplish goals in the conduct of their lives (see Austin, 1975).

It has been shown to be the case, however, that people use different language features more frequently or less frequently for different purposes (see, e.g., Conrad & Biber, 2001). In linguistics, this area of scholarship is sometimes called "English for Special Purposes." Studies in this area often define "registers" of discourse that, essentially, categorizes texts by their situations of use (see, e.g. Biber et al., 1998, chapt. 6). Such linguistic studies are strikingly similar to the study of genres by rhetoricians in that both make use of groupings of discourse and the features of the discourse itself serve as the units of analyses (also see section 2.2, p. 16). A primary purpose of such analyses is to better understand languaging behavior by describing it.

In this study, each genre in the corpora represents a rhetorical situation for which the texts' authors wrote. The idea is to see whether or not particular language features used in the texts vary according to the specific purposes represented by the text genres. As I will explain in chapter 6, it is hoped such findings will lead toward improvements in rhetorical and writing pedagogy, as have resulted in ESL pedagogy from Biber's work.

The text features to be counted in this study are those described by the rhetorical theory promulgated by Kaufer, Ishizaki, Butler & Collins (In Press) and called a "theory of rhetorical priming." This theory attempts to describe usages of everyday English as individual acts of symbolic action. It provides an hierarchy of micro-rhetorical language features that are hypothesized to help account for such actions. The names of the 18 language features in this hierarchy are listed in table 3.2. Each of these

TABLE 3.2: Eighteen categories of rhetorical priming (Kaufer et al., In Press), organized by the three clusters of effect: thought, relations, and description

Thought	Relations	Description
First Person	Reasoning	Word Picture
Inner Thinking	Share Social Ties	Space Interval
Think Positive	Direct Activity	Motion
Think Negative	Interacting	Past Events
Think Ahead	Notifying	Time Interval
Think Back	Linear Guidance	Shifting Event

See chapter 4 for descriptions of each category.

names describes a category of language that is hypothesized to prime a specific micro-rhetorical effect that people use to achieve broader rhetorical purposes.

For example, if my dog is missing I might approach a passerby for help in my search with either of the following utterances:

(1) Have you seen my lost dog?

(2) Have you seen my lost dog with shaggy, blonde fur?

In choosing the second utterance, I decided to give a more complete picture of the dog, hoping to increase my chance of finding him. To enact this goal, I added descriptive language to help my audience visualize my dog. In other words, I chose to use descriptive language ("shaggy, blonde fur") in hopes of priming a word picture of the specific dog for which I'm looking. I could continue the conversation with one of the following:

(3) He's a German Shepherd.

(4) He's a friendly German Shepherd.

In choosing the latter, I added a touch of positivity to my construction, perhaps hoping this will motivate my audience to reach out rather than be afraid of the dog. Toward this end, I attempted to prime a more positive outlook on the German Shepherd breed by revealing my own characterization of the dog's temperament ("friendly").

In choosing one construction over another in each case, I made linguistic decisions aimed at achieving the goal of finding my dog by getting my interlocutor to help look for him. To help achieve this goal, I primed specific micro-rhetorical effects for my audience. The word "primed" is an important one, chosen deliberately to reflect a communicative attempt without certainty of its conclusive rhetorical effect (cf. Coe, 1990; Richards, 1991b). As the speaker, I can never be certain my audience will visualize my dog or believe he is friendly. Nonetheless, I attempt to increase the chances of this by making the specific linguistic choices I make.[2]

Notice both "think positive" and "word picture" are among the 18 categories of rhetorical priming described by the language theory. All 18 of the categories are described in chapter 4. These are the language features to be counted for the study: the percentages of text devoted to each category serve as the study's response variables.

[2]The theory does not require—and I don't intend to imply—these micro-rhetorical choices are always conscious. Like so much of languaging, the choices would most often be tacit to the native speaker engaged in the moments of situated activity (cf. Wittgenstein, 1958).

3.1.3 Count occurrences of the text features

Biber (1988, pp. 213–217) used a two step process to count the selected linguistic features in his study. Biber's software identified the grammatical category of each word in the corpus using a dictionary and attempted to disambiguate words that could be counted in more than one category (such as "blonde" might be either an adjective or noun) by using a series of increasingly complex computer algorithms. Words that could not be so-categorized were left uncounted or were hand-coded. Biber estimated his tagging software plus his hand-coding achieved coverage of 90% or better of the words in his corpus.

The counting procedure in this study is more automated because the theory of rhetorical priming is most fully instantiated in the coding scheme applied to the corpus by a software application called "DocuScope" (see appendix A and Kaufer et al., 2002). This tagging and visualization software was developed in the English Department at Carnegie Mellon University to attempt to help advanced writing and design students. The software distinguishes the language strings associated with the 18 priming categories from all the rest of the text. DocuScope accomplishes this by matching a large collection of language strings (compiled by researchers over the last 5 years) against the text being tagged (Kaufer et al., 2000).

The theory of rhetorical priming is best operationalized as part of this software as it applies the large coding scheme consistently across the corpus of texts. As DocuScope tags the text of a corpus, it does no analysis. The software only categorizes matched strings according to its catalog and

outputs frequency counts as percentages of text for each of the 18 priming categories. This is similar in functionality to other dictionary-based tagging programs (Stone, 2000; Scolari, 2000; Fellbaum, 1998).

Although experience using this fully-automated approach of counting suggests the software is unlikely to yield a text coverage approaching the 90% Biber's hand-coding achieved, using this automated approach provides for an assessment more focused on the particular language theory. This study casts a narrower net than did Biber's in that it does not attempt to include a wide range of features to increase text coverage. This study only counts occurrences of the categories in the text that could be identified as such by the tagging software.

This automated procedure achieves precise, focussed consistency across all texts at the cost of achieving less text coverage. This tradeoff is acceptable because it helps assure independence between the theory and the corpus, avoiding the pitfalls of basing the theory itself on the co-occurrences of features suggested or derived from the corpus texts (see Perfetti, 1998). The software achieved a coverage of approximately a third of the text in the corpora ($\bar{x} = 31.49, SD = 5.21$). I provide extended examples of tagged texts in chapter 4 so you can see occurrences of the tagged features in text samples drawn from the Brown corpus itself.

The summary statistics resulting from the counting procedure for the Brown corpus are shown in table 3.3 and the full counting results from the corpora are available for download from Carnegie Mellon's StatLib (Vlachos, 2001).[3] The summary table provides the mean, median, and

[3]The text tagging was accomplished using a compiled tagging catalog dated 16 Decem-

standard deviation scores for each of the 18 language categories. The mean score indicates the average percentage of text devoted to the category. The median, like the mean, is a measure of central tendency. The median indicates the middle score in the group of scores. The median is less susceptible to outlying scores than is the mean score. The standard deviation score indicates the variability within the corpus of texts on the priming category. In other words, low mean and median scores indicate that, in general, a low percentage of strings in the texts are devoted to priming the category of micro-rhetorical effect. A low standard deviation indicates the texts in the group have similar percentages of strings devoted to the category.

The table also provides the Anderson–Darling normality test statistics. These statistics indicate the normality of the distribution of scores. The p-value in the table's right-hand column indicates the probability the distribution of scores meets the Anderson–Darling test's assumption of normality (at the 95% confidence level). As you can see, the majority of the variables are not normally distributed.[4] This result is encouraging because the statistics shown in the table include *all* of the genres of the corpus. Since the study's hypothesis is that the text's genre affects the us-

ber 2002. I preserved this catalog on CD-ROM, along with tagged versions of the corpora. All the statistical analyses presented in this study were carried out using MiniTab (Ryan et al., 2000) software running on a Macintosh G3 iBook (Mac OS X) in a Virtual PC environment (Windows 98). See Carver (1999) for information about using MiniTab to carry out such investigations.

[4]A low p-value on the test indicates the null hypothesis should not be rejected. This means there is sufficient evidence to reject the assumption of normality of the samples.

TABLE 3.3: Descriptive statistics for the Brown Corpus ($n = 500$ texts)

				Anderson–Darling[a]	
Variable	Mean	Median	SD	A^2	p
First Person	0.564	0.220	0.941	58.302	$\ll$0.001
Inner Thinking	2.691	2.635	0.870	1.394	0.001
Think Positive	0.765	0.700	0.411	5.464	<0.001
Think Negative	1.237	1.150	0.657	3.674	<0.001
Think Ahead	1.203	1.140	0.476	4.659	<0.001
Think Back	0.586	0.520	0.343	7.073	<0.001
Reasoning	2.561	2.540	0.837	0.694	0.069
Share Soc Ties	2.035	1.835	1.010	8.958	<0.001
Direct Activity	0.284	0.220	0.273	31.918	$\ll$0.001
Interacting	0.635	0.350	0.771	32.690	$\ll$0.001
Notifying	2.672	2.660	0.662	0.784	0.042
Linear Guidance	4.767	4.350	2.085	5.767	<0.001
Word Picture	5.354	5.160	2.219	1.899	<0.001
Space Interval	1.330	1.110	0.779	15.891	$\ll$0.001
Motion	0.501	0.400	0.392	14.051	$\ll$0.001
Past Events	2.159	1.975	1.066	6.675	<0.001
Time Interval	1.322	1.230	0.536	11.497	$\ll$0.001
Shifting Events	0.822	0.790	0.323	2.129	<0.001

[a] 95% Confidence level for the A–D normality test statistics

age of the priming categories, one would expect a non-normal distribution across the entire corpus.

In appendix C (p. 236) the same summary statistics are provided by text genre for each of the variables. This is potentially more useful because it shows the different usages of the variables in the different genres. However, we'd like to know whether the differences in scores are statistically significant or if they are simply the result of random fluctuation in language usage. To find this out, analyses via linear models would be

helpful (such as an analysis of variance, ANOVA).

Unfortunately, the Anderson–Darling test statistics indicate that none of the response variables are normally distributed across all of the genres. Furthermore, the spreads of the scores are not homogenous across genres, meaning some scores vary widely within a genre while others do not. Both of these findings indicate undesirable violations of the assumptions that would be necessary for carrying out valid analyses via linear models.

3.1.4 Perform multivariate factor analysis

Factor Analysis (FA) is a multivariate statistical procedure that can help overcome such difficulties and provide insight into co-occurrences of the language features by mathematically reducing the response variables into a new set of linear combinations, called "factors." The factors are more likely to satisfy the assumptions needed for analyses via linear models because they are combinations of all of the variables in a study.

Essentially, factor analysis helps researchers make sense of response variables by identifying patterns of simple correlations among them (Kachigan, 1991; Bryant & Yarnold, 1995). In this study, FA is used to make sense of the co-occurrences of micro-rhetorical priming features, making it simpler to perceive any patterns of the features found within the corpus texts (as is well-described by Biber, 1988, chapter 5). Each of the resulting factors is a mathematical construct that accounts for the main sources of variation in the correlations among the variables. The hope is that such factors will be meaningful when interpreted and can be associated with the text genres. Such interpreted factors would bridge a divide between

the micro-rhetorical language theory and the macro-level rhetorical plans represented by the text genres.

It may be easier to understand the goals of FA if we consider its use in regard to a more familiar set of response variables. As an example, let's suppose I compiled a list of characteristics about the dogs staying at a kennel during a month. After collecting the data, I might use FA to summarize the quantitative relationships among variables that distinguish the dogs from one another in qualitatively important ways. Factor analysis might reveal, for example, the dogs' length and weight are correlated with one another and are in a bi-polar relationship with the dogs' pitch, frequency and duration of sustained barking. This finding would indicate, essentially, that smaller dogs in the kennel had a high-pitched yap and used it more often and longer than bigger dogs.

What does this help me accomplish? It lets me reduce the number of variables to which I need to pay attention. Rather than ask dog owners how much their dogs bark (a subjective evaluation) before reserving the dog a space in the kennel, I might use the information from FA to determine the mix of large and small dogs my kennel could accommodate at any given time while still meeting the city's strict sound ordinances. In other words, FA would allow me to attend to fewer dog characteristics by finding correlations among several pertinent ones. There would likely be other correlations among the data (for example, between the amount of food consumed and the volume of defecation), but FA helps the researcher discover the qualitatively important ones.

For the language study, I used FA to derive a set of factors from roughly

TABLE 3.4: Significant[a] factor loading scores for the five factors extracted from the Brown corpus

Factor 1		Factor 2	
Motion	0.816	Reasoning	0.771
Space Interval	0.790	Inner Thinking	0.753
Word Picture	0.715	Think Negative	0.692
Past Events	0.528	Linear Guidance	0.378
Linear Guidance	0.392	Share Soc Ties	0.368
Interacting	0.351		
(Share Soc Ties)	-0.597	(Time Interval)	-0.347
(Time Interval)	-0.591		
(Notifying)	-0.565		
Factor 3		**Factor 4**	
Think Back	0.752	(Think Ahead)	-0.802
Past Events	0.648	(Direct Activity)	-0.753
Linear Guidance	0.624	(Interacting)	-0.363
Shifting Events	0.589		
(Notifying)	-0.442		
(Direct Activity)	-0.337		
Factor 5			
(Think Positive)	-0.746		
(First Person)	-0.678		
(Interacting)	-0.415		
(Shifting Events)	-0.399		

[a] The critical value for significance is $|0.334|$ (239 texts with overall $\alpha = 0.01$). The full factors (including non-significant variables) are found at table 3.6

half of the Brown corpus texts and then tested the factors' efficacy on the rest of the corpus data. This procedure involves two approaches: the first is called "exploratory factor analysis" and is used to derive a set of factors from data; the second is called "confirmatory factor analysis" and is used to test a set of previously derived (or hypothesized) factors on data. In our kennel example, the first approach would be used to discover from data that small dogs bark more than big dogs. The second approach would be used to help confirm this mathematical finding by applying the factors to a new set of data to see whether the expected relationships hold.

The description of this statistical process is lengthy, may not interest many readers, and, therefore, has been relegated to a later discussion (section 3.3.1, p. 62). The remainder of this section provides an overview of the quantitative methods used and presents their results. My qualitative interpretations of these results are found in chapter 5.

To derive the factors from the data, I used the principal components method of extraction (Stevens, 2002, chapter 11) to extract five-factors from roughly half of the Brown texts. These texts were chosen randomly and the text selection was blocked by genre to ensure each genre was represented, except for "miscellaneous" texts for the reasons described previously. I used only half the Brown texts in this procedure because I needed to reserve the other half for use in confirming the efficacy of the extracted factors, a later step.

I used a mathematical rotation (varimax) to help improve the interpretability of the derived factors. This procedure doesn't change the variance explained by the factors, but it improves their interpretability by

maximizing the high and low loading scores within each factor. This procedure resulted in a set of five uncorrelated factors, shown in table 3.4. As you can see, each factor is comprised of the study's response variables.

Notice each variable in each factor has an associated value, known as its loading score. The differing magnitudes of these loading scores indicate the relative importance of each variable within the factor, regardless of score's sign. This means, for example, in factor one the variables "motion" (with a score 0.816) and "Share Social Ties" (with a score −0.597) are more important variables than those with scores nearer to zero (like "interacting," for example).

The factor loadings can be considered somewhat analogous to the amounts of ingredients in a cooking recipe. Each factor contains different amounts of its ingredients, which, in this case, are the priming categories of the rhetorical theory. Each variable's loading score specifies the amount that goes into the factor. As you'd expect, larger loading scores (in either the positive or negative direction) for a variable indicate the variable has a relatively greater role in contributing to the variance described by the factor. The variables with loading scores nearer to zero have less effect on the factor. In other words, the relative importance of a variable is represented by its distance from zero in either direction. As you can see from table 3.4, each factor has different amounts of the ingredients, suggesting each highlights different relationships among the variables in the underlying data.

The goal of exploratory factor analysis is, ultimately, the meaningful interpretation of the factors. Toward this end, not all of the response

TABLE 3.5: Two presentations of the rotated factor loading scores for the dog kennel example

(a) Extracted loading scores

Variable	Factor 1	Factor 2
Length of Stay	0.198	-0.046
Dog Length	0.706	0.253
Dog Weight	0.647	0.509
Dog Age	-0.047	0.018
Food Consumed	0.013	0.692
Water Consumed	0.004	0.439
Bark Pitch	-0.566	-0.020
Bark Duration (avg)	-0.723	0.137
Bark Volume (avg)	0.166	0.122
Bark Frequency	-0.821	0.064
Defecation Frequency	-0.079	-0.071
Defecation Volume	0.119	0.771

(b) Significant loading scores

Factor 1		Factor 2	
Dog Length	0.706	Defecation Volume	0.771
Dog Weight	0.647	Food Consumed	0.692
		Dog Weight	0.509
		Water Consumed	0.439
(Bark Frequency)	-0.821		
(Bark Duration)	-0.723		
(Bark Pitch)	-0.566		

variables in a study are considered important in each factor. That is, the interpretation of every factor doesn't involve every variable: it only involves the variables with larger loading scores. To better understand why some of the variables in the study are not significant in some factors, let's return to our contrived dog kennel example.

Suppose an exploratory factor analysis of the kennel data resulted in the two factors shown in table 3.5(a). Both of these factors include *all* of the variables in the kennel study, but notice the magnitude of some of the variables' loading scores is near zero. This means these variables play a small role in that factor. There are rules of thumb and statistical procedures for calculating how far away from zero a loading score needs to be for the variable to be considered significant. This value is called the "critical value" and, as I explain in section 3.3.1, its calculation is based upon the number of data points in the study and the overall level of significance desired.

For the language study presented in this book, the critical value for determining whether or not a loading score is significant is 0.334. Again, the sign of the loading isn't a consideration here—as long as the absolute value of the loading score is greater than the critical value the variable is considered significant (and, therefore, may be a salient feature of that factor). Notice the factor loading scores in table 3.4 are all greater than this critical value.[5]

If we use this same critical value for our contrived dog kennel example, we could reformat table 3.5(a) to eliminate the non-significant variables. That is, we'd drop any variable with an absolute value less than 0.334. Furthermore, we could also group the variables together according to their signs because variables with like signs are correlated with one another (e.g. a dog's length and its weight are correlated). This makes the

[5]The 0.334 critical value provides an overall confidence level of $\alpha = 0.01$ (twin-tailed) based on the 239 Brown texts used in extracting the factors (see Stevens, 2002, p. 394).

table somewhat easier to read and interpret, as you can see in table 3.5(b) in which I've dropped the non-significant variables and grouped together variables with like signs.

Factor 1 in this kennel example is bi-polar, meaning it has both positive and negative significant loadings (Kachigan, 1991, p. 254–56). This doesn't mean some dogs have somehow learned to bark negatively. The bi-polar loading scores simply indicate that dogs with higher scores on the positively-loaded variables in the study tended to have have lower scores on the negatively-loaded variables. Likewise, dogs with low scores on the positive variables had high scores on the negative ones. Essentially this means big dogs tend to bark less than small dogs and vice versa. The groups of positive and negative variables occur in a complementary distribution. Notice factor two is not bi-polar. That is, the variables with like signs on the factor correlate with one another, but there is no complementary set of negatively-loaded variables.

Leaving the kennel behind, table 3.4 (p. 46) is the end result of the FA extraction procedure from half of the texts of the Brown corpus. The table is formatted to aid understanding of the factors and includes only the significant variables in each factor.

Once the factors had been extracted, I then calculated an individual factor score for each text in the corpora. Having these scores enabled comparisons of the different genres of text to one another, discussed below. Essentially, the texts' factor scores could provide insights into how genre affects the usage of the micro-rhetorical priming categories making up each factor.

Before getting to that step of analysis, however, it was important to test the derived factors against a new set of data to see whether or not the relationships held up across the corpora. Biber (1988) did not use such a procedure in his language study, but the procedure, called "confirmatory factor analysis," is recommended by Stevens (2002, pp. 411–53) and is deemed crucial by Benjamin B. Lahey, a researcher at the University of Chicago who uses FA in his studies of childhood psychological disorders (personal communication, 12/26/01).

The procedure involves using Pearson correlation and chi-square tests to compare group means of factor scores to one another. The procedures involved in these analyses are discussed in detail in section 3.3.1 (p. 71). The Pearson correlation and the chi-square results provide statistical evidence for the efficacy and durability of the factors, essentially confirming the factors shown in table 3.4 at the $\alpha < 0.05$ confidence level.

These quantitative results indicate the factors initially extracted from one half of the Brown corpus (randomly selected, blocking by genre) also describe the correlations among the variables in the other half of the Brown and in the Frown corpus. This is remarkable evidence for the importance of these factors because it demonstrates their usages remained fairly consistent according to text genre between 1961 and the early 1990s when the texts of the Frown corpus were written.

Previous rhetorical studies have demonstrated marked genre change over time (e.g., Bazerman, 1994b; Berkenkotter & Huckin, 1995). Even the brief comparisons of the tobacco ads of the 1940s and 1950s that introduced this study (chapter 1) showed how the usages of some rhetorical

priming categories is subject to change.

Between the composition of the texts in the Brown and the Frown corpora, we might expect to find changes in several of the genres. For example, we might expect the "science fiction" genre to change from the Brown corpus's 1961 sci-fi texts (written before the lunar landing) to science-fiction of the 1990s from the Frown corpus (post-cyberspace, pre-nanotechnology). Or we might expect changes in the "press reportage" genre, which would be comparing pre- and post-Watergate collections of reporters' writing samples. Of course any consistent changes in the authors' usages of the rhetorical priming categories would affect the genre's typical scores on the factors. And, indeed, we do find changes in some of the genres' scores, as are described elsewhere in this study (see chapters 5 & 6).

Yet the confirmatory factor analysis shows the derived factors remained stable across the corpora, suggesting that the factors may mark quite fundamental decisions by authors and could be valuable for helping to understand rhetorical choice in text composition, a finding further supported by the qualitative evaluation provided in chapter 5.

3.1.5 Determine whether the factors are affected by text genre

A key question underlying this study is whether the genre of a text affects the usage of the micro-rhetorical priming features that make up the factors. I chose to answer the question, in part, using statistical analyses of variance. These tests quantify the significance of the differences be-

tween the genres of texts in the corpus. These procedures, like the other statistics, are detailed more fully in section 3.3.2 (p. 75).

These procedures are useful because they compare the scores of the genres on the factors, take account of each genre's variability, and provide an indication of how confident we can be that the differences between the factor scores is due to the texts' genre and is not the result of random variation among the texts.

The results of these analyses indicate the usage of all five of the factors is affected by the text genre (in both the Brown and Frown corpora) at the $\alpha < 0.001$ level of significance. This means there is a 99.9% (or better) probability the text's factor score is related to the genre and is not a random occurrence. This engenders great confidence in saying the genre of a text affects the usage of those rhetorical priming categories that make up the five factors (shown in table 3.4 on p. 46).

To better understand how a text's genre affects the usage of the factors, I also made comparisons of the genres to one another using another statistical process called "Tukey's method" (Tukey, 1949). This method compares each individual genre to all the others simultaneously (known as making "pairwise comparisons"). The results of this test are provided in the figures beginning on page 77 and show that many of the genres' factor scores are separated from other genres with 95.0% confidence. For example, Tukey's method indicates the factor 1 scores of "learned" texts and newspaper editorials differ significantly. That is, authors of academic texts tended to use the rhetorical priming categories of factor 1 differently than the authors of the newspaper editorials did.

These quantitative results answer a key question of the study in the affirmative: the rhetorical priming features are, indeed, affected by the text genre in detectable, statistically significant ways. This indicates that a fuller understanding of the factors may lead to a fuller understanding of some of the linguistic behaviors of text production.

The question remains, however, whether or not the derived factors can be interpreted meaningfully. That is, we still need to ascertain whether or not the factors are qualitatively important for language researchers—and, potentially, writers and writing teachers—to understand.

3.1.6 Interpret the factors as language dimensions

The interpretation of extracted factors is a difficult process that involves comparing the factor loadings with the texts and the genres themselves. The goal of this qualitative interpretation is to explain what each factor indicates about the usage of the salient rhetorical language features in the different genres of the corpus. Successfully interpreting the factors would provide for possible insight into the meso-level rhetorical strategies employed by the authors of the texts.

Interpreting the factors of the dog kennel example looks easy because dogs and dog characteristics are quite familiar to most people and because the example was contrived to allow the factors to pop out meaningfully. Real data is not so easily interpreted, as is discussed by Biber (1988) and Bryant & Yarnold (1995). Its interpretation requires careful understanding of the response variables behind the factors and the data underlying the factors.

The following sections are intended to help you understand how I arrived at the interpretations of the factors. Section 3.2 of this chapter describes the 15 genres of the Brown and Frown corpora, giving you an overview of the generalized, high-level rhetorical plans instantiated by the texts of each genre. Chapter 4 describes rhetorical priming theory in detail and provides examples of the priming categories from the Brown corpus texts. The interpretation of each of the factors as a language dimension is then detailed in chapter 5, beginning on page 127.

3.2 Description of the genres of the corpora

This section provides a brief description of each of the text genres that make up the corpora. The 15 genres are listed in table 3.1 (p. 33). The compilers of the Brown corpus (participants at a conference at Brown University in February 1963) drew up the list of genres and subdivided them into two groups: "informative prose" (genres A–J'); and "imaginative prose" (genres K–R). The corpus excludes drama and poetry (other than short quotations of verse). The corpus compilers also excluded text samples with more than 50% dialogue (Francis & Kucera, 1979, p. 2).

The placement of genres into the two groups, "informative" and "imaginative," will strike us as odd today. Sensibilities about what comprises "imaginative" work has surely changed since 1963 (when the division of the corpus was conceived) to include, at least, biography and memoir. Nonetheless, I've repeated the division here and in chapter 5 because the labels may have affected the placement of individual texts in the differ-

ent genres and because the compilers of the Frown corpus kept the same genres and divisions in creating their parallel corpus (Hundt et al., 1999). Referring to the groups of genres also proves convenient in explaining the interpretations of some of the language dimensions (in chapter 5). The division of text into the two groups is not considered in the statistical procedures of this study.

3.2.1 The "Informative Prose" Genres

A: Press Reportage

The texts in this genre make up the bulk of newspapers and news magazines. In relatively short space these texts describe all manner of subjects for readers, including local happenings, sports, finance, and world events. They cover a multitude of topics from presidential veto threats to football game recaps to world-wide diplomatic and scientific efforts.

B: Press Editorials

Newspapers and news weeklies generally include several essays that provide the opinion of the publisher and of paid columnists or prominent citizens. These essays cover topics the publication deems would be of interest or importance to its readers. The pieces in this genre might suggest, for example, how the reader should vote in an upcoming election or they might describe an upcoming event in flattering or unflattering terms in order to sway opinion about the event. Some authors try to bring historical understandings to their readers in order to help them see a contem-

porary happening in a different light. The genre also includes published letters to the editor; often these are written in response to a previous article or editorial essay.

C: Press Reviews

Criticism of the arts is a staple of popular journalism. The texts of this genre critique television programs, operas, recently-opened plays, movies and recently-published books. The reviews that make up this genre seem intended to give the reader insight on the work being reviewed, letting the reader know what the reviewer thought and why he or she thought it.

D: Religion

This genre consists of published articles, religious tracts, and theological essays, some published in specialized journals and others as books. The texts cover a broad range of human behavioral topics, including social and religious practices. These texts usually place the behavior or result of the behavior into scriptural context or they recommend alternative courses of action from a well-defined perspective.

E: Skills and Hobbies

These texts address the broad range of interests and pursuits in which Americans engage from cooking to travel to farming to managing businesses to tracking baseball statistics to buying and refurbishing boats. The texts, drawn mostly from periodicals, seem oriented toward people engaged or potentially engaged in the pursuit that is discussed. Many of

the texts seem intended to offer practical advice on whatever the topic of the text.

F: Popular Lore

Texts written on specific topics for a generalized audience seem to make up the bulk of this genre. The texts mostly introduce a topic to the audience with varying degrees of detail. Some of the eclectic topics covered by these books and periodicals include parapsychology, part-time farming, health remedies of the pioneers, single parenthood, tooth-straightening, and therapy by witchcraft. The texts of the genre were drawn from both books (specialized on the topic) and periodicals.

G: Belles Lettres, Biography, Memoirs

Equally divided between book excerpts and articles from periodicals, this genre includes biographies of notable contemporary and historic people, published letters on a variety of aesthetic and literary topics, a range of published memoirs, and essays on the high arts. The texts mostly focus on happenings and events (especially in biographies) or on emotions or thoughts resulting from happenings (such as a recent visit to a museum or a recent read of an authors' collected works) or life events.

H: Miscellaneous

This genre is essentially a catch-all category. It includes mostly government documents (studies, reports, handbooks) and reports from foundations and industry groups. It also includes excerpts from a college catalog

(Carleton College) and an annual financial report from an unidentified company.

J: Learned

The texts of this genre are academic books and articles published across the spectrum of disciplines: natural sciences, medicine, mathematics, social and behavior sciences, political science, law, education, humanities, and engineering. Most of the texts seem intended to present the results of some research the author(s) completed or have engaged. Other texts seem intended to address the community of scholars to an issue deemed important by the author(s).

3.2.2 The "Imaginative Prose" Genres

The "imaginative" genres are comprised of excerpts from novels and also from short stories selected from anthologies, periodical collections (e.g. *The Virginia Quarterly*) and magazines (e.g. *The Atlantic Monthly*).

K: General Fiction

The texts of this genre appear to have been classified as "general" because they didn't fit into the other, more topical "imaginative prose" genres (described below). The topics of the "general fiction" texts are widely varied.

L: Mystery and Detective Fiction

This genre is comprised of texts in which some crime (usually violent) has been committed and is under investigation. The reader of these texts usu-

ally follows along with the main investigator, sometimes knowing more than the character and sometimes learning of clues and happenings along with the main investigator.

M: Science Fiction

It appears the texts of this genre were classified as "science fiction" because they involve considerations of space or time travel or of the future. The texts seem devoted to explaining the world drawn by the text, but some also involve intricate plots with multiple characters.

N: Adventure and Western

The texts of this genre cover a broad range of happenings, but usually they appear set in the 19th century American west with titles such as *Rifle for Rent.* Some of the texts, however, are set elsewhere and center around main characters learning to flourish (or at least survive) in difficult environments containing challenging obstacles. Like the "mystery and detective fiction" genre, the reader is usually positioned to follow a main character.

P: Romance and Love Story

The texts of this genre are usually written from the point of view of a character engaging physically and emotionally with characters of the opposite gender in a wide variety of public and intimate settings.

R: Humor

This genre includes texts that seem written to be funny, exposing juxtapositions of real life with idealized or commonly-held conceptions of life. Sometimes farcical, the texts vary in setting and topic.

3.3 Statistical methods

This section provides more detail about two of the statistical methods used in this study, factor analysis and analysis of variance.

3.3.1 Factor analysis

This section complements section 3.1.4 (p. 44) by further describing the procedures used to derive the factors and test their efficacy.

This procedure involved two related approaches: the first is called "exploratory factor analysis" and was used to derive the set of factors from the corpus data; the second is called "confirmatory factor analysis" and was used to test the set of derived factors. Refer to Stevens (2002, chapter 11) for a detailed description of the mathematical processes involved. The qualitative interpretations of the results of these processes are provided in chapter 5.

To accomplish the two processes, I first divided the Brown corpus in two, reserving roughly half the texts for use in each process. To split the corpus, I randomly selected half the texts in each genre. For genres with an odd number of texts, I rounded up. This procedure resulted in two

groups of Brown texts, randomly drawn but blocked by text genre with half a consisting of $n_a = 254$ texts and half b with $n_b = 246$ texts.

Exploratory factor analysis

I used the principal components method of extraction to estimate the number of factors and the factorability of the correlation matrices among the texts in half a of the corpus. I excluded texts in the "miscellaneous" genre, for reasons described previously.

Initially I extracted 18 factors (the number of response variables) with the intent of keeping only factors with eigenvalues above 1.0, a common rule of thumb suggested by Kachigan (1991, p. 246). The eigenvalue is a measure of the amount of variance accounted for by a factor. An eigenvalue of 1.0 corresponds to an amount of variance corresponding to one of the original variables. Keeping the factors with eigenvalues greater than 1.0 ensures each factor accounts for more variance than a single variable would. Thus, those factors are determined to contribute significantly (statistically) to explaining the relationships among the response variables of the study. Five of the extracted factors met this criterion.

This number of factors coincides with a second procedure for estimating the number of useful factors to be extracted (see also Stevens, 2002, pp. 389–90). This second procedure uses a scree curve of the factors' incremental variances. The curve is examined to determine where it first flattens, indicating that factors subsequent to that point explain less significant variance in the underlying data (see figure 3.1) .

Since both methods of determining the number of factors roughly agree,

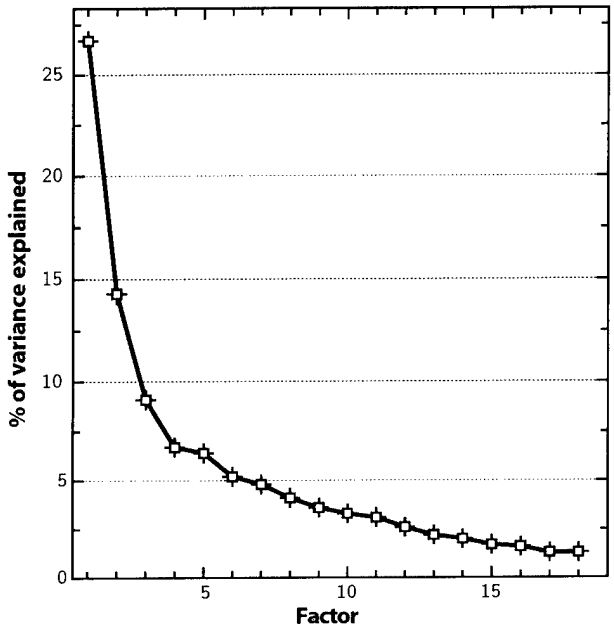


FIGURE 3.1: Variance accounted for by 18 unrotated factors extracted through the principal components method. Factors were extracted from the response variables of randomized half *a* of the Brown corpus

I repeated the extraction procedure, this time requesting a 5-factor solution and using a varimax rotation to improve the chances of viewing the factors from an interpretable angle (Stevens, 2002, pp. 391–93). A varimax rotation redefines the extracted factors, sharpening the distinctions between them. Such a rotation aids in the interpretation of factors. The rotation eliminates as many medium-sized loadings as possible, but it does not change the total variance explained by the factors and it maintains the original factors' orthogonal structure. This procedure results in factors that are comprised of response variables with high and low factor

loadings and that are uncorrelated. Because of this, rotated factors often better-highlight distinctions among texts that may then be meaningfully interpreted.

Biber (1988, pp. 81–85) also chooses to rotate the extracted factors in his study, but he uses a promax rotation, rather than varimax rotation. Biber chooses promax because, he explains, "Varimax maintains orthogonal structure, requiring the assumption that the factors are uncorrelated, while promax permits oblique structure, that is, it permits minor correlations among the factors" (p. 85). Biber recommends using an oblique rotation for language data, explaining his recommendation in a footnote in which he hypothesizes the following:

> ...oblique solutions might be generally preferable in studies of language use and acquisition, since it is unlikely that orthogonal, uncorrelated factors actually occur as components of the communication process. That is, from a theoretical perspective, all aspects of language use appear to be interrelated to at least some extent, and thus there is no reason to expect mathematically uncorrelated factors representing those aspects. (Biber, 1988, p. 85n2)

However, later in his methodology Biber (1988, pp. 93–97) effectively undermines this hypothesis when he eliminates some of the language features that were initially included in more than one of his rotated factors. He provides his reason for doing so:

> ...to assure the experimental independence of the factor scores,

> each feature was included in the computation of only one factor score. Thus, each linguistic feature is included in the factor score of the factor on which it has the highest loading (in terms of absolute magnitude, ignoring plus or minus sign). (Biber, 1988, p. 93)

Furthermore, Stevens (2002) suggests correlated factors may be "more reasonable to assume in most cases" (p. 392), but he cites statistical investigations by Pedhazur and Schmelkin (1991) in which they write,

> From the perspective of construct validation, the decision whether to rotate factors orthogonally or obliquely reflects one's conception regarding the structure of the construct under consideration. It boils down to the question: Are aspects of a postulated multidimensional construct inter-correlated? The answer to this question is relegated to the status of an assumption when an orthogonal rotation is employed... (qtd. in Stevens, 2002, p. 392)

In consultation with professors in the Carnegie Mellon Statistics and English departments (Vlachos and Kaufer, personal communication, 2/2/03), it was determined that the maintenance of uncorrelated factors using the varimax rotation would be a reasonable assumption and would maintain a degree of experimental independence of the factors that Biber attempted to achieve post hoc by modifying his factor structures manually.

As I mentioned previously, each factor developed in this study represents a combination of the rhetorical priming categories. These combina-

TABLE 3.6: Rotated factor loading scores (varimax rotation)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
First Person	0.198	0.046	0.227	0.070	-0.678
Inner Thinking	-0.200	0.753	-0.007	-0.158	-0.161
Think Positive	-0.147	0.108	-0.164	0.054	-0.746
Think Negative	0.083	0.692	0.042	0.090	0.021
Think Ahead	-0.284	0.039	0.076	-0.802	0.035
Think Back	-0.066	0.120	0.752	0.010	0.089
Reasoning	-0.279	0.771	0.073	-0.027	-0.075
Share Soc Ties	-0.597	0.368	-0.329	-0.152	0.005
Direct Activity	0.187	-0.040	-0.337	-0.753	0.094
Interacting	0.351	0.309	0.131	-0.363	-0.415
Notifying	-0.565	0.055	-0.442	0.043	0.188
Linear Guidance	0.392	0.378	0.624	0.151	-0.163
Word Picture	0.715	-0.309	0.081	-0.003	-0.137
Space Interval	0.790	-0.170	0.247	-0.019	-0.005
Motion	0.816	-0.035	0.013	-0.008	0.012
Past Events	0.528	-0.073	0.648	0.223	0.066
Time Interval	-0.591	-0.347	0.078	-0.106	0.034
Shifting Events	0.221	-0.183	0.589	0.006	-0.399
Eigenvalue	3.737	2.332	2.322	1.492	1.488
% of Variance	0.208	0.130	0.129	0.083	0.082

tions are described by series of numbers—called factor loading scores—with one loading score assigned to each priming category in each factor. Table 3.6 provides the factor loading scores for the five rotated factors in this study.

The goal of exploratory factor analysis is the meaningful interpretation of the factors. Toward this end, not all of the variables are considered salient in each factor. Variables with lower loading scores are dropped from the factors. The remaining, statistically significant vari-

ables are considered during the interpretation of the factors (Stevens, 2002, pp. 393–94; Kachigan, 1991, pp. 251–55).

Several rules of thumb exist for deciding on the magnitude of loading necessary for a variable to be considered significant. Probably the most common (and the one followed by Biber, 1988, p. 87) is to accept all loadings with a magnitude greater than $|0.35|$ as being significant enough for interpretation. However, Stevens, p. 394 cites Monte Carlo evaluations by Cliff and Hamburger (1967) indicating the necessity of taking sample size into account by calculating the critical value for simple correlation at $\alpha = 0.01$ (a two-tailed test). This procedure provides for a more stringent control of the overall significance level and reduces the probability of false rejections accordingly.

Given the number of texts used to extract the factors ($n_a = 239$, which is the randomly drawn half a of the Brown corpus minus the 15 texts in the "miscellaneous" genre), the critical value for factor loadings for this study is $CV = 0.334$. As before, the sign of the loading does not affect its significance. Therefore, to be considered statistically significant and, thus, be considered during the interpretation of the factor, the magnitude of a variable's factor loading has to be greater than the absolute value of the critical value. In other words, variables with an absolute value of loading score greater than 0.334 are accepted as statistically significant.

Taken as a whole, the process I've just described comprises exploratory factor analysis, with table 3.4 (p. 46) as its result. The table is formatted to aid understanding of the factors by including only the significant variables in each and grouping the variables together by their sign. These

resulting factors are interpreted in chapter 5, but several other statistical procedures are run using the factors to help determine their explanatory power.

Calculating factor scores for texts

Using the derived factors, one can calculate a factor score for each individual text in the corpora (Biber, 1988, pp. 93–97). The resulting scores can then be used for comparisons of the texts (or groups of texts) to one another and other statistical procedures (Kachigan, 1991; Stevens, 2002). In this study, the text's factor scores are needed to understand whether or not text genre affects the texts' scores on the factors. In other words, these scores will allow us to see whether or not the genre of the text affects an author's usage of the priming categories making up each factor.

Calculating factor scores for texts begins with a table of factor coefficients, generated during the exploratory factor analysis procedure. The factor coefficients are mathematically related to the factor loading scores and are provided in table 3.7.

The factor score for a text is calculated by multiplying the text's standardized score on each response variable by the variable's factor coefficient, and then summing these products. Standardized scores, often called "Z-scores," are calculated for texts in a group by subtracting the mean group score from the text's score and then dividing the result by the standard deviation of the group. Basically, the standardized score indicates how many standard deviations and in which direction away from zero each text in a group resides on that response variable. Standardizing

TABLE 3.7: Rotated factor coefficients (varimax rotation)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
First Person	-0.021	-0.053	0.015	0.037	-0.471
Inner Thinking	-0.001	0.312	-0.005	-0.073	-0.038
Think Positive	-0.068	-0.048	-0.155	0.084	-0.580
Think Negative	0.099	0.346	-0.043	0.090	0.113
Think Ahead	-0.109	-0.050	0.210	-0.580	0.036
Think Back	-0.155	0.017	0.458	-0.089	0.152
Reasoning	-0.038	0.324	0.049	0.009	0.028
Share Soc Ties	-0.112	0.125	-0.077	-0.045	-0.025
Direct Activity	0.142	-0.001	-0.125	-0.497	0.065
Interacting	0.105	0.109	-0.004	-0.254	-0.226
Notifying	-0.098	0.017	-0.138	0.090	0.060
Linear Guidance	0.047	0.173	0.232	0.044	0.014
Word Picture	0.201	-0.092	-0.089	-0.027	-0.079
Space Interval	0.222	-0.011	0.000	-0.057	0.061
Motion	0.294	0.073	-0.151	-0.012	0.070
Past Events	0.060	0.002	0.259	0.066	0.141
Time Interval	-0.268	-0.237	0.203	-0.101	-0.055
Shifting Events	-0.088	-0.150	0.275	-0.074	-0.250

the scores centers the group of texts on zero while preserving the differences between the texts in the group. This makes comparisons across variables meaningful and makes it feasible to combine variables, as in the factors.

For example, the first factor score of Brown text A01 is calculated by summing the products of the the variables' coefficients (the left column of table 3.7) and the text's standardized variable scores. For text A01, this process yields a factor 1 score $= -0.581$. This procedure was repeated for each text in the corpora for each of the factors.

TABLE 3.8: Correlation matrix for factors derived from randomly selected half of the Brown corpus (half a) and applied to the other half

	Factor 1_a	Factor 2_a	Factor 3_a	Factor 4_a	Factor 5_a
Factor 1_b	0.895	0.293	0.647	0.138	-0.496
	<0.001	0.289	0.009	0.624	0.060
Factor 2_b	0.213	0.859	0.435	0.198	-0.542
	0.446	<0.001	0.105	0.479	0.037
Factor 3_b	0.680	0.366	0.973	0.071	-0.239
	0.005	0.180	<0.001	0.802	0.391
Factor 4_b	0.027	0.507	-0.026	0.809	-0.455
	0.924	0.054	0.926	<0.001	0.088
Factor 5_b	-0.420	-0.355	-0.159	-0.184	0.760
	0.119	0.194	0.571	0.512	0.001

Cells contain the Pearson correlation and p-value ($df = 13$)

Confirmatory factor analysis

To confirm the derived factors, I moved to another approach called "confirmatory factor analysis" (Stevens, 2002, pp. 411–53). For each of the two halves of the Brown corpus (divided randomly, as described on p. 62), I calculated the mean factor score for each of the 15 genres. The summary statistics for the genres are provided in appendix D.

To compare the two halves of the corpus and, thus, "confirm" that the a priori factors fit the data of half b of the corpus, I examined simple Pearson product moment correlations, comparing the mean factor scores of each genre to the same scores from the other two groups of texts (Bryant & Yarnold, 1995, pp. 110–11). The Pearson correlation coefficients represent standardized covariances (each item is fixed to one), accounting for

important differences in group variability that can be obscured by correlations. Essentially, the Pearson coefficient indicates the degree of linear relationship between two variables. The resulting values range from $+1$ to -1, indicating a positive or negative correlation. The resulting p-value represents the likelihood the correlation coefficient is zero, which would indicate an absence of correlation. The Pearson results are shown in table 3.8. These indicate each of the five factors derived from half a produce correlated responses when applied to the texts in half b.

I repeated this procedure for the genres of the Frown corpus, with the results shown in table 3.9. The results indicate the factors derived from half a of the Brown corpus produce correlated results when applied to the texts making up the genres of the Frown corpus. Remember that the two corpora are "parallel," meaning the compilers of the Frown corpus used genres identical to the Brown corpus, except the texts were published some 30 years later (Hundt et al., 1999).

To quantify the overall model fit of the five factors to the different groups of texts, I examined the chi-square fit statistics, comparing the genres of half a of the Brown corpus to the other two groups (half b and the Frown texts). The chi-square statistic combines a measure of the strength of a relationship between values with information about the sample size to provide an overall indication of whether or not the relationship achieves statistical significance.

The chi-square statistic is, in general, a robust measure of the goodness-of-fit and is recommended (along with other, more specialized, statistical methods) for use in confirmatory factor analysis. However, the statistic

TABLE 3.9: Correlation matrix for factors derived from randomly selected half of the Brown corpus (half a) and applied to the Frown corpus

	Factor 1_a	Factor 2_a	Factor 3_a	Factor 4_a	Factor 5_a
Factor 1_f	0.916	0.223	0.598	0.102	-0.482
	<0.001	0.425	0.018	0.717	0.069
Factor 2_f	0.151	0.884	0.556	0.125	-0.413
	0.590	<0.001	0.031	0.657	0.126
Factor 3_f	0.564	0.291	0.899	0.109	-0.394
	0.029	0.293	<0.001	0.699	0.146
Factor 4_f	-0.010	0.254	-0.223	0.630	0.007
	0.973	0.361	0.423	0.012	0.982
Factor 5_f	-0.629	-0.341	-0.396	-0.225	0.805
	0.012	0.214	0.144	0.421	<0.001

Cells contain the Pearson correlation and p-value ($df = 13$)

TABLE 3.10: The number of genres with mean factor scores in each quartile by corpus

	Brown (half a)[a]				Brown (half b)				Frown Corpus			
Factor	1q	2q	3q	4q	1q	2q	3q	4q	1q	2q	3q	4q
1	1	6	3	5	1	5	4	5	1	6	1	7
2	2	4	7	2	2	3	8	2	0	5	8	2
3	2	4	4	5	2	5	3	5	1	6	4	4
4	4	6	4	1	2	6	6	1	5	6	3	1
5	3	9	3	0	4	7	4	0	3	7	4	1

[a] Counts for Brown (half a) were used as the expected values in the chi-square test (results shown in table 3.11)

TABLE 3.11: Chi-square fit statistics for the factors

		Brown (half b)		Frown Corpus	
Factor	df	χ^2	p	χ^2	p
1	3	0.500	0.919	2.133	0.545
2	3	0.393	0.942	2.393	0.495
3	3	0.500	0.919	1.700	0.637
4	3	2.000	0.572	0.500	0.919
5	2[a]	1.111	0.426	0.778	0.678

[a] For factor 5 $df = 2$ because the genres fell into only three of the quartiles in the half a of the Brown corpus

requires categorical variables and has been found to be less reliable when some categories have low expected counts (Stevens, 2002). Since the factor scores in this study are continuous measurement variables, I used the median and inner-quartile range statistics of each group to construct a contingency table, grouping the 15 mean genre scores of each corpus by the appropriate quartile. Thus each of the mean genre scores was placed into one of the four quartiles for each factor, as indicated in table 3.10.

The results of the chi-square tests are provided in table 3.11. The chi-square's p-value represents the likelihood the test's assumed relationship between the model fits should be rejected. In other words, a small chi-square statistic and a large p-value reflects a good fit between the mean factor scores of the text genres in one group and the second. As you can see, these analyses indicate all of the factors derived from half a of the Brown corpus fit the data from half b of the Brown corpus and from the Frown corpus (at the $\alpha < 0.05$ confidence level).

This finding coincides with the Pearson results, suggesting we may confidently apply the extracted factors to the other groups of texts. As I discuss in chapter 5, the interpreted factors (called "language dimensions") are also found to be qualitatively useful in understanding the broader rhetorical underpinnings of the genres of both the Brown and the Frown corpora. That chapter provides qualitative evidence that further buttresses the confidence in the efficacy and durability of these five derived factors.

3.3.2 Analyses of Variance

The following procedures make use of *all* the texts of the Brown and Frown corpora, including the texts of the "miscellaneous" genre, which were excluded from the exploratory factor analysis procedure. The factor score for each text was recalculated using the entire corpora's mean scores to calculate the normalized response variable scores (refer to section 3.3.1 for details of calculating factor scores for a text).

To compare the genres on all of the factors simultaneously, one-way multivariate analysis of variance (MANOVA) was carried out. MANOVA indicates whether or not the genres of texts differ significantly on the factor scores. I chose to consider the most conventional statistic, the Wilks' Lambda test (Wilks, 1932), which yielded $F(70, 2294) = 18.73$, $p < 0.001$ for the Brown corpus and $F(70, 2294) = 16.09$, $p < 0.001$ for the Frown texts.

When the sample sizes of groups are nearly equal, the one-way ANOVA is not overly sensitive to inequality of variances. However, when one

TABLE 3.12: Analysis of variance results testing for an effect of genre on factor score

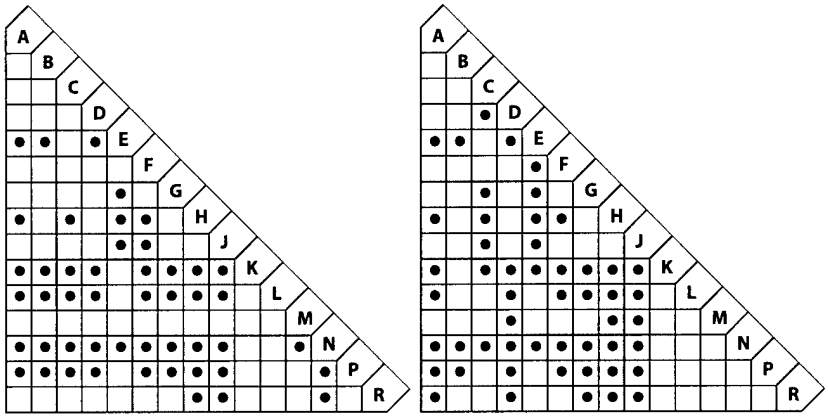
	Brown Corpus		Frown Corpus	
Factor	F	p	F	p
1	26.19	$\ll$0.001	31.71	$\ll$0.001
2	13.15	<0.001	7.64	<0.001
3	22.35	$\ll$0.001	18.25	<0.001
4	7.63	<0.001	7.59	<0.001
5	7.86	<0.001	13.41	<0.001

$n = 500$ texts per corpus; $df = 14, 485$

group is much larger than another (as in this study, with genres ranging from 6 to 80 texts) the calculated F statistic may be dominated by the variances in the larger sample. I evaluated this possibility by randomly sampling 6 Brown corpus texts from each genre 1,000 times and calculating the Wilks' Lambda score each time. This Monte Carlo procedure provided a distribution of Wilks' Lambda scores and the variability for a balanced design (with $n = 6$ text samples per genre). The procedure resulted in a mean Lambda that was significant ($F(70, 2294) = 4.541$, $p < 0.001$) with a standard deviation $= 0.615$ for the 1,000 iterations.

The significant result on the MANOVA procedure does not mean that each genre differs significantly from all the others on each of the five factors. To test for this possibility I carried out a univariate analyses of variance (ANOVA) for each factor. All five factors were found to be significantly affect by text genre in each corpus, as is shown in table 3.12.

For each of the factors I also used Tukey's method (Tukey, 1949) to make pairwise comparisons between the genres of texts. This comparison

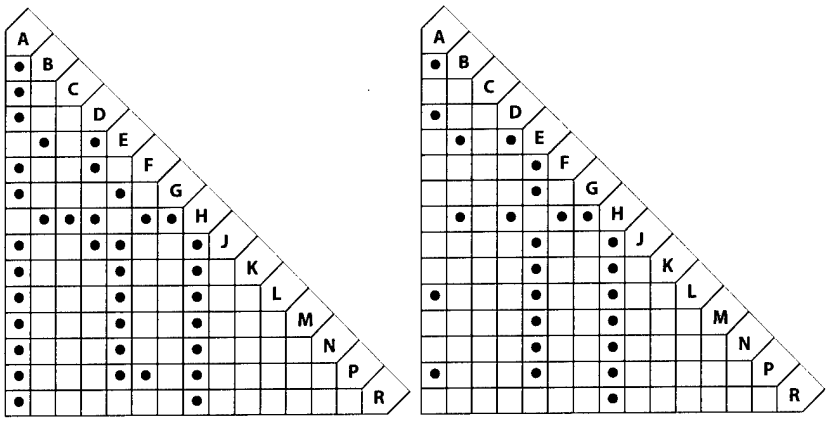


(a) Brown Corpus genres (b) Frown Corpus genres

FIGURE 3.2: Tukey all-pairwise comparison results for factor 1 scores. Marked cells indicate significant difference (at $\alpha < 0.05$ family error rate) between genres

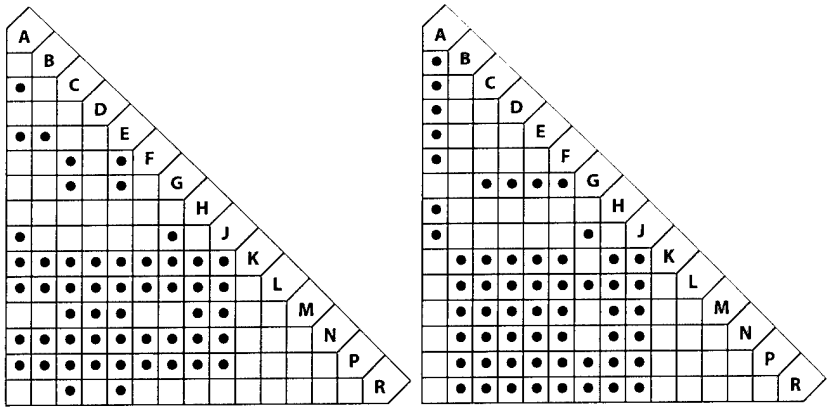
method provides an indication of whether a pair of genres—e.g., *science fiction* and *humor*—differ significantly by constructing confidence intervals for all the pairwise differences between the mean factor scores for the genres. In other words, this test indicates whether each pair of groups differs on the factor being considered.

Tukey's method showed significant differences (at $\alpha < 0.05$ or better) between many of the genres on each of the factors, as is shown in the comparison charts, beginning with figure 3.2(a). From these results it is clear that the usages of rhetorical priming language strings making up the factors are significantly affected by the genre of the text. This is big news, but the question remains whether or not the extracted factors can be interpreted meaningfully, a question addressed in chapter 5.



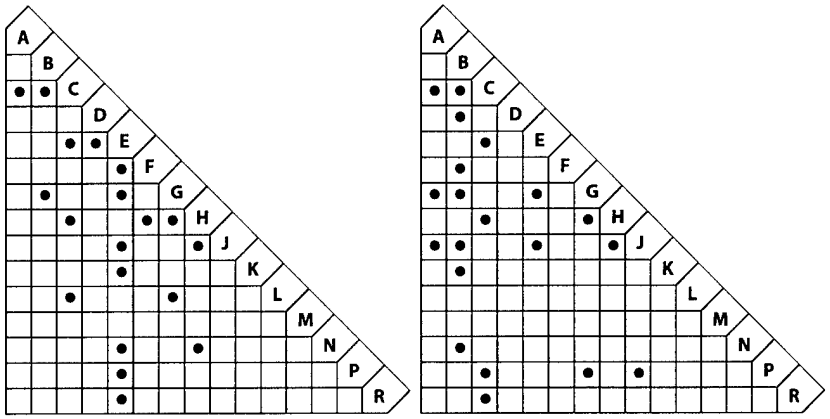
(a) Brown Corpus genres (b) Frown Corpus genres

FIGURE 3.3: Tukey all-pairwise comparison results for factor 2 scores. Marked cells indicate significant difference (at $\alpha < 0.05$ family error rate) between genres



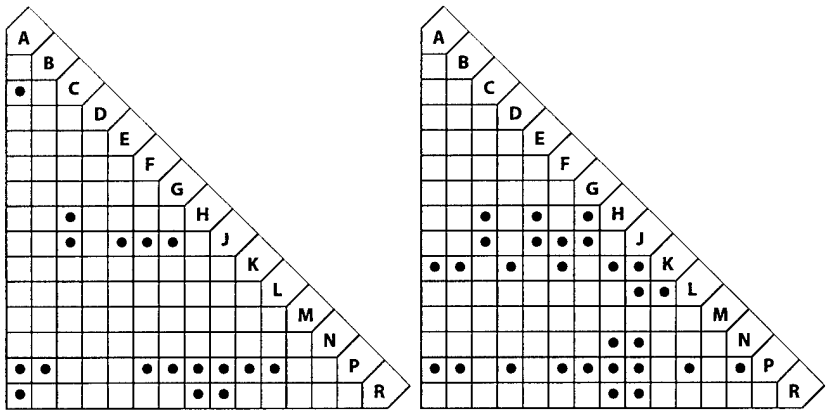
(a) Brown Corpus genres (b) Frown Corpus genres

FIGURE 3.4: Tukey all-pairwise comparison results for factor 3 scores. Marked cells indicate significant difference (at $\alpha < 0.05$ family error rate) between genres



(a) Brown Corpus genres (b) Frown Corpus genres

FIGURE 3.5: Tukey all-pairwise comparison results for factor 4 scores. Marked cells indicate significant difference (at $\alpha < 0.05$ family error rate) between genres



(a) Brown Corpus genres (b) Frown Corpus genres

FIGURE 3.6: Tukey all-pairwise comparison results for factor 5 scores. Marked cells indicate significant difference (at $\alpha < 0.05$ family error rate) between genres

4 Rhetorical Priming Theory

This chapter provides a further introduction to the theory of rhetorical priming and a detailed description of each response variable in this study. The theory is defined fully in Kaufer, Ishizaki, Butler & Collins (In Press).

This chapter supplements the current study by providing an example of the high-level distinctions about writing and rhetorical plans the theory is capable of describing. The chapter also provides examples of the micro-rhetorical decisions the theory hypothesizes writers control. Finally, this chapter demonstrates how this theory is instantiated as a coding scheme used to tag the corpora for this study.

4.1 Achieving goals through different high-level rhetorical plans

It's not particularly surprising that different rhetorical plans might be used to achieve any particular goal. For example, if the chemical DDT hadn't been banned from widespread use in the U.S., writers might want to argue DDT should be banned. Writers in such a situation could choose

from a myriad of options for designing an argument. Below are three possible plans a writer might choose to enact:

Plan 1 *Write an explicit argument that will provide the reader with the main reasons DDT ought to be banned.*

Plan 2 *Write a paper outlining the writer's opinions of DDT and what she thinks about its continued use.*

Plan 3 *Write a description of DDT's effects (or potential future effects) on the environment.*

Arguments based on these three rhetorical plans would lead to quite different experiences for a reader: the first would feel like a research paper, leading the reader to consider the presented evidence critically; the second would feel like an opinion piece, leading the reader to consider how much credence to pay the author and his or her opinion; the third would feel like a narrative, leading the reader to visualize the world described in the paper and to compare it to the one the reader lives in and knows. These striking differences in reading experience would be created linguistically—the plan chosen will lead the writer to choose language that creates the particular experiences the reader will have in the text.

All three of these plans could potentially lead to effective papers arguing against DDT. The idea behind rhetorical priming theory is that it may be useful to understand the way micro-level language choices combine to create, or at least "prime," these different experiences for readers. In

other words, the theory attempts to categorize the linguistic choices authors make that lead readers to have *particular* experiences with texts.

Kaufer et al. (In Press) argue writers design these experiences for their readers as they write. Writers aren't necessarily using the features explicitly, but are simply trying to achieve their aims or trying to create a "feel" in their texts. Furthermore, writers can't predict with absolute assurance that their texts will achieve the "feel" the authors hoped for, let alone achieving the overall goal for a text. Communication is, after all, a stochastic process to some degree, especially communication intended to persuade.

The rhetorical theory, along with the tagging and visualization software, DocuScope, were designed for use in writing courses to let students see some of their micro-level rhetorical usages explicitly. The intent was to help students dialog about and gain better control of these potentially important linguistic choices, as I discuss more fully in the following sections.

4.2 Enacting rhetorical plans through cumulative, micro-rhetorical decisions

To help demonstrate the point of the language theory and the software, let's consider the subtle linguistic shifts that accompany changes in high-level rhetorical plan.

For example, read the following sentences from potential anti-DDT arguments and consider into which plan they would fit appropriately:

(5) I think there are fewer birds in my neighborhood now than when I moved here a decade ago—their beautiful singing no longer announces the dawn.

(6) During the 10-year period of the study bird populations decreased an average of 35.8% in many mixed land-usage areas (farming and suburban), with populations of wrens and species of song birds found to have dropped by as much as 62.0% in some areas.

Both sentences provide evidence to support the writer's assertions; however, the kind of evidence used in the sentences differs. Sentence (5) might work in an opinion piece or as a personal narrative of environmental changes (plans 2 or 3), but it probably wouldn't work as well in a formal research report (plan 1). Likewise, sentence (6) might throw off a reader engrossed in a personal narrative or naturalistic description of the countryside (although see Bryson, 2001, for an example of how authors can combine evidence for use in hybrid types of arguments).

Not only does the kind of evidence in the sentences differ, but so do the linguistic choices made to introduce the evidence. Evidently the writers made linguistic decisions appropriate to the kinds of evidence they chose to provide. Consider how incongruous it would seem to come across the following sentence:

(7) I think there are an average of 35.8% fewer birds in mixed-use land areas becasue they seem to sing less than before.

This sentence would not fit well in any of the plans because the phrases "I think" and "seem" don't appropriately signal the type of exacting quan-

titative evidence the sentence also provides.

Furthermore, even within a single rhetorical plan, writers use subtle language choices to help achieve rhetorical purpose. For example, the original (8) and revised (9) sentences below were written by a property owner to the manager of the rental property:

(8) Please ask if the tenants would be willing to move out early to accommodate renters who want to move in during the summer.

(9) Please ask if the tenants would be willing to move out early to accommodate <u>any possible</u> renters who <u>might</u> want to move in during the summer.

There is nothing particularly wrong with sentence (8) to cause its revision, but the writer achieved a change in rhetorical effect by adding the underlined words.

The rhetorical difference between these sentences was achieved through fairly subtle language change. The first sentence implies renters have been lined up and are ready to move in during the summer. In the second sentence, the renters seem more tentative and uncertain, suggesting it is the reader (the property manager) who would know about potential renters, not the writer of the letter. The writer in sentence (9) primed indications of inner thinking. This means "the renters" moved from being explicit, tangible, described people in (8) to being figures of imagination or potentiality in (9). They've gone from being in the world to being imagined in the writer's mind. They're not renters, but possible renters who might (and hopefully will someday) exist. The revised sentence conveys

a slightly more tentative way for the manager to approach the current tenants and suggests the manager should take action to look for and line up future renters.

Of course there are other ways a writer could achieve this impression. It's quite likely a writer would combine this sentence with others in the letter to help clarify the relationship and to make the direction to the manager more careful (or perhaps the direction is already clear from the broader context in which the letter is written). Nonetheless, by making the specific language changes to the first sentence, the author revised it for rhetorical purpose—a purpose probably designed to help make the overall point of the letter.

Kaufer et al. (In Press) argue experienced writers have control of such micro-rhetorical shifts, subtlely manipulating language here and there in attempts to achieve different impressions for readers (see also Kaufer & Butler, 2000). This is what is meant by the cumulative rhetorical effect of language choice: no one choice necessarily makes a strong impression, but cumulatively the choices lead to particular impressions for readers. Historically, writers have attained this control through years of reading and writing practice—both in the school setting and beyond it. To try to help writing students begin to understand these cumulative language effects more explicitly, Kaufer and Butler, along with designers Suguru and Kerry Ishizaki, created DocuScope.

4.3 Finding priming effects in text

The two previous sections have described the theoretical usage of micro-rhetorical language choices for effecting high-level rhetorical plans. This section describes how a catalog of such effects is used as coding scheme for software, DocuScope, that finds and tags the priming effects in a text.

In the examples presented so far, the language changes have been rather subtle, effecting small rhetorical differences by adding touches of one effect or another to a construction. Creating an extensive coding scheme to capture such rhetorical effects has been a difficult, on-going challenge. The challenge is made more difficult because identical words and even phrases might be part of extremely different micro-rhetorical experiences for a reader. For example, consider the different effects of the word "smeared" in the following sentences:

(10) John smeared butter on his toast.

(11) John smeared his opponent.

"Smeared" is priming quite different experiences for the readers of these two sentences. In sentence (10) *smeared* indicates a mundane, everyday physical motion, but in (11) it primes a strong negative affect (there's likely no one who would choose to be "smeared" by an opponent).

A coding scheme or computer program that attempted to tag rhetorical function based on the usage of single words would miss the rhetorical difference between these sentences. Some software parsers, looking at rhetorical or grammatical structure (e.g. Marcu, 2000), would have the flexibility needed to parse the sentences and find the grammatical func-

tionality of the usages, but such systems provide focus on categories that are blind to rhetorical effect.

The coding scheme used for this study was designed specifically to identify language strings authors might use to prime distinct effects for readers (listed in section 4.5). For example, word strings such as those given below would appear in the coding scheme, differentiating two of the priming effects involving the word *smeared:*

Motion
smeared butter
smeared margarine
smeared tanning oil
smeared jam
smeared [article] butter
smeared [article] margarine
etc.

Think Negative
smeared [pronoun] opponent
smeared [pronoun] enemy
smeared [pronoun] opposition
smeared [article] opponent
smeared [article] enemy
smeared [article] opposition
etc.

Over the past five years, a large catalog of such language strings (over 450 million) has been compiled by researchers classifying each string into 1 of 18 categories of rhetorical priming (Kaufer et al., In Press). The tagging software, DocuScope, makes use of this catalog to provide an estimate of the experiences a text will make available for readers.

In highlighting the strings associated with these 18 categories, DocuScope makes available a portion of the information we believe readers attempt to gain through close readings of texts. The software (and the language theory behind it) by no means make available all the information from a close reading: they capture little of style, for example, and do

not attempt to interpret language for deep semantics or cultural significance. Nor can the software account for rhetorical structures that depend on discontinuous elements across sentences and paragraphs. The strings captured by the software represent what a myopic reader of English with an impoverished understanding of structure, context and culture could find.

Although this "reading" is challenged in these ways, the software finds and tags these strings more consistently than could a human reader. It automatically applies the coding scheme to collections of texts, tagging each text and indicating the quantity of each of the 18 variables found (given as a percentage). This enables careful assessment of the tagged language features in a way similar to the use of other dictionary-based tagging programs and parsers like Wordnet or Diction (Stone, 2000; Fellbaum, 1998; Scolari, 2000).

The quantifications resulting from DocuScope's processing are estimates, tokens, of the theoretical amount of rhetorical priming in the texts. The estimates are just that: the computer only matches strings from its coding scheme and these strings were placed within the scheme by human researchers. Although the researchers have done their best to be complete and consistent, the placement is based on subjective judgments. Such estimates lack perfect intersubjectivity. That is, different writers and different readers might make different decisions about where specific language strings should be placed withing the hierarchy of such strings. Nonetheless, as this study and common experience illustrate, there is value in understanding subjective judgments about language. People commonly

experience this value when they ask someone else to read their texts and value the feedback given them. DocuScope gives perfectly consistent feedback on one aspect of writing, micro-rhetorical priming.

4.4 A hierarchy of micro-rhetorical priming effects

Talbot Taylor (1997) has suggested that in the post-Wittgenstein era, intersubjectivity isn't a necessity (or even a possibility) for communication theories. The language theory and software underlying this study aren't attempts at providing a grand, intersubjective story of communication. Instead, these tools are attempts to help students and teachers get on the same page in regard to how they talk about language usage (Kaufer & Ishizaki, 2000; Collins, 2002). The theory and software were developed to provide an occasion for invention,[1] for interaction, and for reflection in the writing process, all probably good things (cf. Huot, 1990; Flower, 1994; Huot, 1996; Yancey, 1999). The lack of intersubjectivity, far from making the software less valuable in a writing classroom, makes it potentially more valuable by inviting discussion and reflection on the shared artifact (i.e. the tagging and resulting visual tokens generated by DocuScope).

Essentially, the classroom users of DocuScope develop a shared metadiscourse surrounding each text and writing occasion, limited only by the sensitivity of the language theory and consistency of the computer pro-

[1]Bakhtin (1986) in "Speech Genres" points to the transactional nature of language and language occasions as sources of invention.

gram. (Consistency is not the same as intersubjectivity). In other words, DocuScope is an attempt to move toward "linguistic reflexivity" in the type of discourse used by students and teachers surrounding language usage (see Taylor, 1997, chapter 9).

Of course this highlights a distinction between practical and intellectual metadiscourses. In using a tool like DocuScope in a writing classroom, teachers might move past the conundrum of talking about language with students as a theoretical communicational vehicle (intellectual metadiscourse) to an emphasis on exploring social conformity in communicative practice—what rhetoricians and genre theorists have long advocated (not "right/wrong," but "better/worse" for a given rhetorical situation). As Taylor and others advocate, teachers should not wish students to look exclusively at what language theorists claim about communicational understanding or rhetorical strategies, but, instead, to raise questions that lead students to analyze language reflexively—to ask questions about why they have applied this language move here or that type of language there.

In initially developing the theory, Kaufer & Butler (1996, 2000) theorized about the types of metadiscursive practices the features of their theory would need to engender for writers. Recognizing rhetoricians' long-interest in patterns of language usage, Kaufer and Butler began to operationalize their theory into computer software that would consistently tag specific language features. As they did so they were better able to characterize authors' usages of the features and, eventually, they developed a careful hierarchy of such rhetorical features, which is depicted in fig-

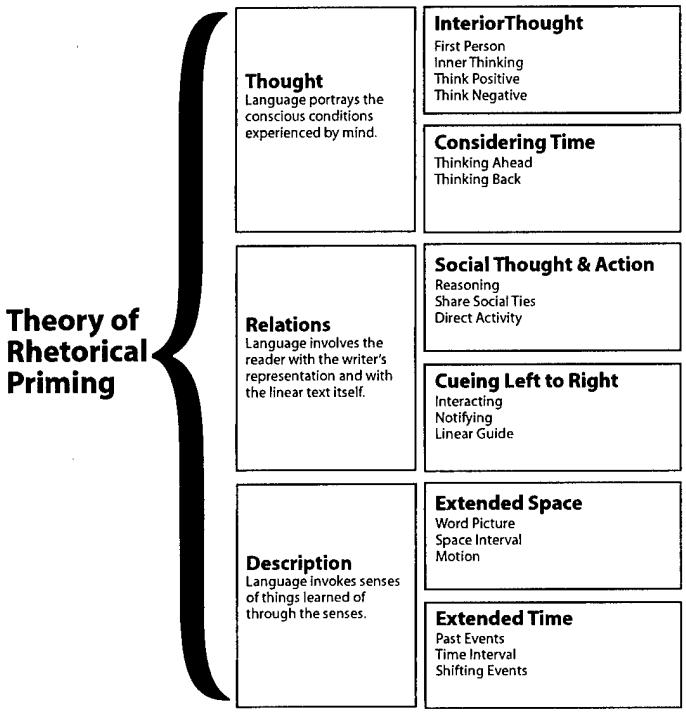


FIGURE 4.1: Eighteen micro-rhetorical priming categories, organized by the three clusters (thought, relations, and description) and six families of effect. (Adapted from Kaufer, Ishizaki, Butler & Collins, In Press, p. XX)

ure 4.1 (p. 91).

As described in the latest book (Kaufer et al., In Press), the hierarchies of features bear some rough overlaps with ideas found in the works of rhetorician I.A. Richards (1991a) and linguist M.A.K. Halliday (1994). The theory reflects Richards' important notion that much of English separates into concepts of inner thought and outer sense. An important determiner for a rhetorical action is whether it is mental action spawning from a mind (as in "I think fewer birds are singing") or outward action, pro-

jecting a descriptive reality (as in "the research team counted 36% fewer *trodlodytes ædon* this year than last").

Based on Richards' ideas, two major divisions (called "clusters") were constructed at the top of the hierarchy, *thought* and *description.* These two clusters of categories, described in sections 4.5.1 and 4.5.3, provide the distinction between language strings priming disclosure of thoughts and those priming the immersion of a reader in spatial and temporal situations. *Thought* and *description,* together, create the combined inner and outer depictions required for priming reader awareness and familiarity with ideas and situations.

The third cluster in the hierarchy of priming effects is described, at least in part, by Halliday's systemic-functional grammar. Halliday's idea is that a fundamental function of language involves what he labels the "interpersonal metafunction," the ability of language to structure interactive relationships with audiences.

The priming categories in this middle cluster, called *relations,* prime the rhetorical effects of information on readers. That is, they prime how an audience is to process the text and how they are to take it. For example, the two sentences below suggest different ways for a reader to take information:

(12) According to Alexander Hamilton's view of federalism, the central government should retain strong control of the banking function.

(13) Alexander Hamilton suggested the current federal banking system, a key part of his federalist view.

In sentence (12) Hamilton is an authority whose view of federalism is to be explained while in (13) Hamilton is an historic actor whose legacy includes views on federalism.

Such subtle orientation effects are central to reader decision-making insofar as getting readers to comprehend depends upon getting and sustaining their attention through a written chain of reasoning, an immersive example, exhortations, and other textual devices. The reader awareness and comprehension of such devices are primed through the use of the various categories of the *relations* cluster, described in section 4.5.2 (p. 106).

Each of the three clusters of effects is further divided into two subcategories —called "families"—that are indicated on the right side of figure 4.1 by the bold labels in the six boxes. DocuScope's visualization functions use these families to choose the color for each of the matched strings of a text. This use of color is designed to benefit users by drawing attention to specific consistencies and variabilities within and among texts.[2]

How well the clusters and families correspond to a broad range of readers and their reading experiences with texts has not been formally evaluated. Students' anecdotal experiences using the software in the writing classroom indicate this hierarchy is adequate for drawing attention to the priming effects and helping students develop a metalanguage about what is important for the writing tasks they are asked to accomplish. In

[2]Since the families of effect are not part of this study, the reader is referred to Kaufer et al. (In Press) for information regarding them. For more discussion of the theorized benefits of text visualization, see the references cited in appendix A.

other words, the hierarchy and its concomitant color scheme seem to help DocuScope users compare their use of the rhetorical priming effects to the effects in the writings of others.

The purpose of the study described in this book is not to evaluate the clusters and families in the hierarchy. This study is an attempt to describe the differences in actual usage of the individual priming effects for specific high-level rhetorical purposes (the genres), which were described in section 3.2 (p. 56).

4.5 Description of the rhetorical priming effects

The string catalog used for this study characterizes over 450 million strings according to 18 categories of rhetorical priming effect. Below is a brief description of the effects, grouped into the three previously-described clusters: *thought, relations,* and *description.* The definitive description of each dimension may be found elsewhere (Kaufer et al., In Press).

I've followed each description with two examples showing <u>underlined</u> strings the software matches to the priming category. The first example following each description is a group of sentences adapted from a recent news article. Each of the sentences was modified to provide more strings matching the priming category. The second, more lengthy example following each description consists of sentences culled from the texts of the Brown corpus. These sentences were not modified from the original texts and the passages are long enough to demonstrate limitations of the tagging catalog, which I discuss below. Each example text was chosen be-

cause it was the highest scoring Brown text on that priming category.

The amount of text underlined in these examples is somewhat misleading visually because only the priming category being exemplified is underlined in each example. This makes for sparse underlining. Naturally, the underlining is more dense when all of the priming categories are shown in the sentences. As described in section 3.1.3, in any given sentence from the corpora in this study about a third of the text would be underlined as matching the categories of rhetorical priming by the software. This more typical density of underlining can be seen in the screen shots included in appendix A.

Also included below is a rough heuristic diagram[3] for each cluster, designed to help distinguish the priming categories within the cluster from one another. Although I have used the basic form and shapes of a flow chart, these diagrams should not be confused with the flow charts used in computer programming: not all of the decision-blocks are answered with a "yes" or "no" and every string in a text will not be classified by the software (thus, the poof-clouds seen in the diagrams). Despite these differences, I've found the diagrams helpful in distinguishing the priming effects from one another.

As I attempted to make clear in section 3.1.3, such coding of text is based upon the best possible research but is limited by the stamina of those involved in compiling the catalog of priming strings. In other words, at this point in history not every existant English string has been included

[3]These heuristics were adapted from a classroom handout by David Kaufer.

in the catalog. Furthermore, some strings that have been included are not perfectly unambiguous, which means they may match improperly in some cases. These limitations suggest that the amount of priming the software captures in a text is different than the theoretical amount of priming in that text. Nonetheless, the absolute consistency of the catalog's application over a large sample of texts (in this study, over two million words) drawn from a variety of authors and sources tends to minimize the effect of these limitations on the study's findings.

4.5.1 Priming effects of the *thought* cluster

The categories of this cluster prime a reader for an awareness of cognitive functioning. Such priming helps the reader perceive a subjective perspective unambiguously. That is, priming from this cluster leaves little doubt about whether the text is providing interior or exterior representations.

First Person

Writers often need to provide the impression of a mind at work. Whether the represented mind is that of the author or a narrator or other character, self-referential language strings are among the ways writers individuate a point of view. The strings in this category prime notions of a single consciousness, separate from others.

(14) I have peaches in my cart and I'm positive they're organic. This makes me feel better about the groceries I choose for my family

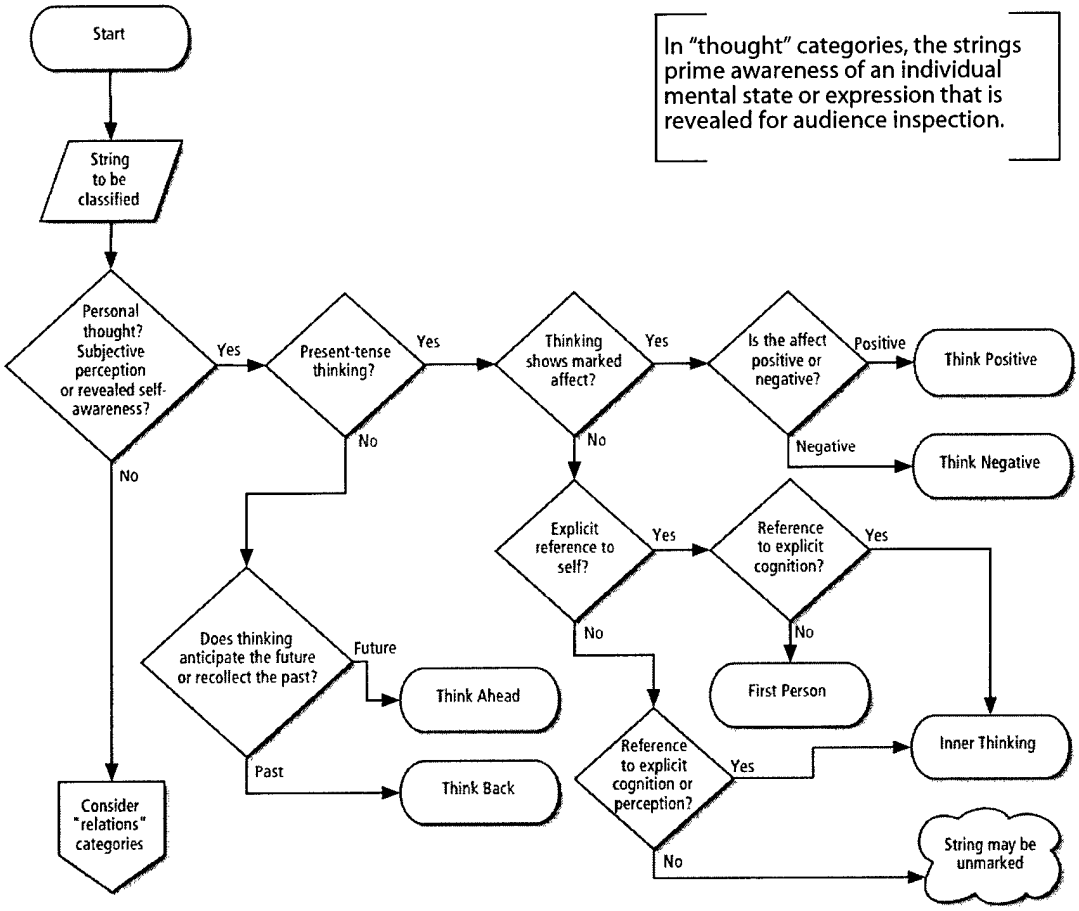


FIGURE 4.2: Heuristic for distinguishing the categories of the *thought* cluster

(and helps me feel better about spending extra for organic produce).

(15) from Brown text N05: Richard Ferber, *Bitter Valley*

a. She was carrying a quirt, and she started to raise it, then let it fall again and dangle from her wrist. "I saw your fire", she said, speaking slowly, making an effort to control her anger. "You could burn down this whole mountainside with a fire that size. It couldn't matter to a fool like you. It would to me".

b. "I'm not advising you", she said. "I'm telling you. That fire's too big. Let it burn down. And make sure it's out when you leave in the morning".

c. "There's some mistake", he said finally. "You're right about the fire. It's bigger than it has to be, though I don't see where it's doing any harm. But you're wrong about the rest of it. I'm not leaving in the morning. Why should I ? I own the place".

d. "You own this place" ? she said, and her tone had softened until it was almost friendly. "You bought it" ?

e. "From a man in St. Louis", Wilson said. "Jake Carwood. Maybe you know him". The girl laughed. "I know him. I ought to. My father ran him off here six years ago". Wilson didn't say anything. He stood watching the girl, wondering what was coming next.

f. "Never mind", she said sternly. "It wouldn't matter to my father, and not to me. I meant what I said about that fire. Be

sure it's out when you leave. That's all. I'll let you go back to doing the dishes now". It was meant to insult him, and didn't quite succeed.

Inner Thinking

Writers don't necessarily require "first person" strings to create impressions of private and particular minds at work. Writers can prime such awareness by using references to cognition and cognitive acts or by revealing subjective perceptions or contingencies, the raw materials of thinking.

(16) Never did we imagine the need to heed the meaning of "organic." Even if we had, there would have been no apparent way to be certain what we were buying was really organic.

(17) from Brown text D10: Huston Smith, "Interfaith Communication: The Contemporary Scene"

a. Massacres attending the partition of India and the establishment of the State of Israel are simply recent grim evidences of the hostility such divisions can engender. The words of Cardinal Newman come forcibly to mind: " Oh how we hate one another for the love of God"! The source of this paradox is not difficult to identify.

b. It lies in institutions. Institutions require structure, form, and definition, and these in turn entail differentiation and exclusion. A completely amorphous institution would be a contradiction in terms; to escape this fate, it must rule some things

out. For every criterion which defines what something is, at the same time proclaims—implicitly if not openly—what that something is not.

c. Some persons are so sensitive to this truth as to propose that we do away with institutions altogether; in the present context this amounts to the advice that while being religious may have a certain justification, we ought to dispense with churches. The suggestion is naive. Man is at once a gregarious animal and a form-creating being.

d. Having once committed himself to an ideal which he considers worthwhile, he inevitably creates forms for its expression and institutions for its continuance.

e. To propose that men be religious without having religious institutions is like proposing that they be learned without having schools. Both eventualities are possible logically, but practically they are impossible.

Think Positive

Writers prime positive feelings in texts by using language strings marked by "feel good" words, words representing emotional affect that would be desired by people.

(18) Recent laws have made it easier to know what you put in your grocery basket, a welcome improvement.

(19) from Brown text C02: Various authors, *Christian Science Monitor*

a. As a matter of fact, this latter approach has already been tried, and with pleasing results. A few years ago a "Timex All-Star Show" offered a broad range of styles, ranging from Lionel Hampton's big band to the free-wheeling Dukes of Dixieland. An enthusiastic audience confirmed the "live" character of the hour, and provided the interaction between musician and hearer which almost always seems to improve the quality of performance.

b. About that same time John Crosby's TV series on the popular arts proved again that giving jazz ample breathing space is one of the most sensible things a producer can do. In an hour remembered for its almost rudderless movement, a score of jazz luminaries went before the cameras for lengthy periods.

c. However well chosen and cleverly arranged, such memorabilia unfortunately amounted to more of an interruption than an auxiliary to the evening's main business, which (considering the talent at hand) should probably have been the gathering of fresh samples of the Chicago style.

d. Still, the network's willingness to experiment in this musical field is to be commended, and future essays happily anticipated.

Think Negative

Writers prime distressful feelings to allow readers insight on negative views held by the writer, narrator, or character in the written work. When writers combine stings of negative affect with neutral descriptions, readers are signaled that the writer disapproves of the described entity. The language strings in this category prime feelings people recognize as negative for the most part.

(20) It was those abuses that produced the restraints that inhibit inspection practices overseas and helped damage the already poor credibility of the bureaucracy.

(21) from Brown text D06: Various authors, *Tracts published by American Tract Society*

a. I have, within the past fifty years, come out of all uncertainty into a faith which is a dominating conviction of the Truth and about which I have not a shadow of doubt.

b. SATELLITES, SPUTNIKS, ROCKETS, BALLOONS; what next? Our necks are stiff from gazing at the wonders of outer space, which have captured the imagination of the American public. Cape Canaveral's achievements thunder forth from the radio, television, and newspaper.

c. While we are filling outer space with scientific successes, for many the "inner" space of their soul is an aching void. Over $200,000,000 is paid yearly to the 80,000 full-time fortune-

tellers in the United States by fearful mankind who want to "know" what the future holds! Delinquency, juvenile and adult, is at an all-time high!

d. Further proof? Read your daily newspaper! Unfortunately, in our rush to beat the Russians, we have forgotten these truth-packed words of Jesus Christ: "What shall it profit a man, if he shall gain the whole world [that includes outer space], and lose his own soul? Or what shall a man give in exchange for his soul"? (Mark 8:36, 37). Gaining outer space and losing "inner" space is bad business according to God's standards.

Thinking Ahead

Writers use this category of priming when they need to project into the future. This category primes readers' awareness that they are reading about an anticipated, possibly tentative, future, not a present or past.

(22) With advent of the upcoming web-based reporting procedure, governments will get into more careful oversight because no bureaucrat will want to be blamed for missing a problem as they hope to make the food supply safer.

(23) from Brown text H01: U.S. Gov't, *Handbook of Federal Aids to Communities*

a. Production specialists in SBA regional offices are available to help individual small business concerns with technical production problems. Guidance and advice are available on new prod-

uct research and development; new product potential; processing methods; product and market developments; new industrial uses for raw, semi-processed and waste, materials; and industrial uses for agricultural products.

b. The property sales assistance program is designed to assist small business concerns that may wish to buy property offered for sale by the Federal Government.

c. Under this program, property sales specialists in the Small Business Administration regional offices help small business concerns to locate Federal property for sale and insure that small firms have the opportunity to bid competitively for surplus personal and real property and certain natural resources, including timber from the national forests.

d. The Administration maintains a productive facilities inventory of small business industrial concerns that have voluntarily registered. It is kept in each Regional office for the small firms within the region. Purpose of this inventory is to include all eligible productive facilities in SBA's facilities register so that the small business concerns may have an opportunity to avail themselves of the services authorized by the Congress in establishing the Small Business Administration.

e. A proposed loan must be for sound purposes or sufficiently secured so as to assure a reasonable chance of repayment. The record of past earnings and prospects for the future must in-

dicate it has the ability to repay the loan out of current and anticipated income.

Thinking Back

At times writers need to let readers know a text is recounting the past and recollecting what had been an assumed or experienced actuality that is now receding into memory. A writer primes such retrospective understandings for readers by using strings from this category.

(24) The legislation has made it easier and may have prevented a problem already. The old law was to have kept shoppers forever guessing about the "organic" label.

(25) from Brown text F30: Fred Birmingham, *The Ivy league Today*

a. There is much to be said for such a college—and Dartmouth men have been accused of saying it too often and too loudly. Their affection for their college home has even caused President Dickey to comment on this "place loyalty" as something rather specially Hanoverian.

b. The legislators decided to "liberate" Dartmouth and entered into a tug-o'-war with the college trustees over the control of classrooms, faculty, and chapel. For a time there were two factions on the campus fighting for possession of the student body. The struggle was resolved in 1819 in the Supreme Court in one of the most intriguing cases in our judicial history.

c. Dartmouth is numerically still a small college today, with approximately twenty-nine hundred undergraduates. But it has achieved a cross-section of students from almost all the states, and two-thirds of its undergraduates come from outside New England.

d. Dartmouth is moving closer to the others in the Ivy group. It is still, however, the junior member of the League, if not in years at least in the catching up it has had to do. It has not been a well-known school for any part of the span the other Ivies have enjoyed.

e. However much football has been over-emphasized, the public likes to measure its collegiate favorites by the scoreboard, so, while Yale need never give its record a thought again since outscoring its opponents 694 to 0 in the season of 1888, Dartmouth had to wait until its championship team of 1925 for national recognition. It has come on with a rush in more significant areas.

4.5.2 Priming effects of the *relations* cluster

The categories of this cluster prime readers' understandings of text processing goals. Such understandings allow readers to more efficiently navigate through text and helps orient them regarding the way the writer has intended them to take representations stemming from the text. As indicated by the two decision diamonds at the left of figure 4.3, the cluster

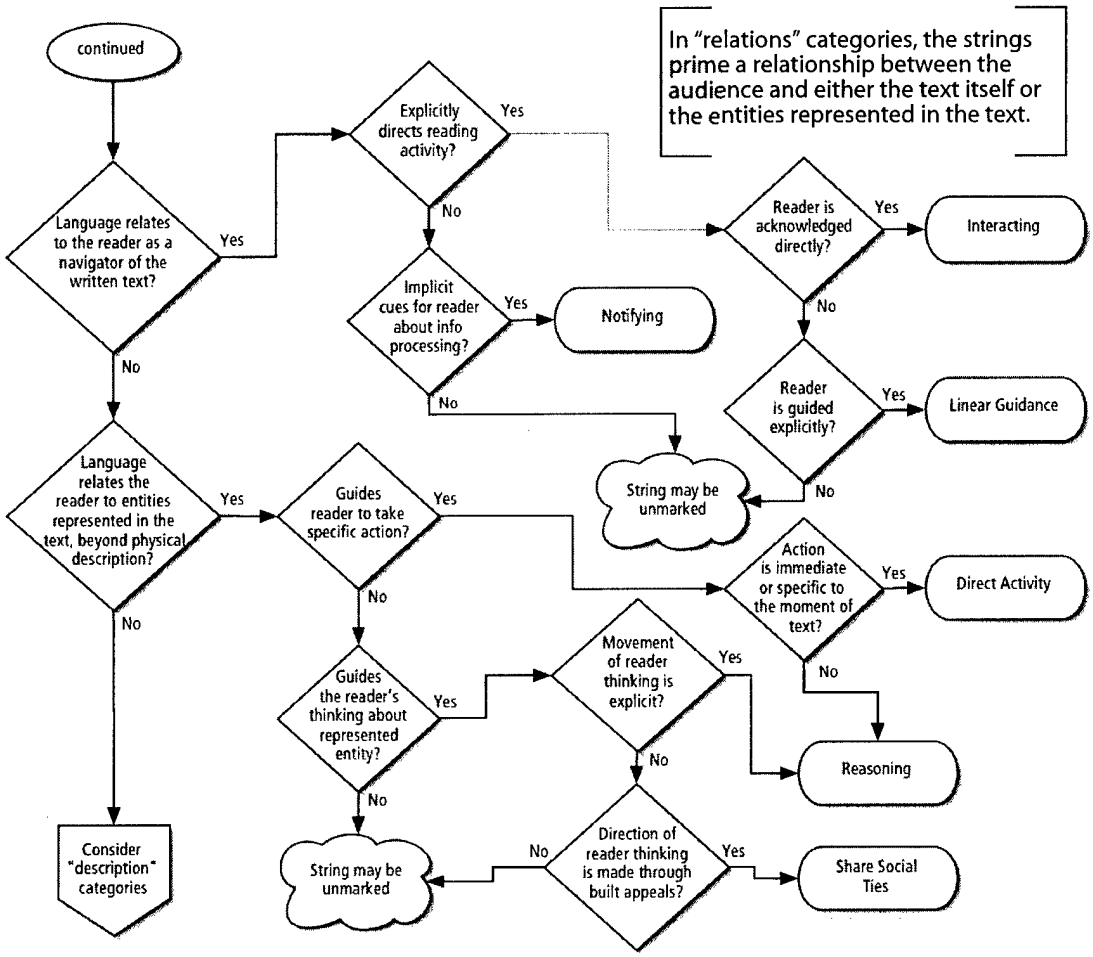


FIGURE 4.3: Heuristic for categories of *relations*

captures two overlapping areas of priming: strings priming approaches to the text itself and strings priming approaches to the entities represented through the text.

Reasoning

Using "reasoning" priming strings, writers guide the reader as a thinking being who, in following the linear text also follows a path of reasoning that is likely not linear. Writers use the strings of this category to prime awareness of the reasoning paths they intend readers to follow or, at least, see. This priming may be the opening means of bidding for assertions or refutations of a path's conclusion. This category is marked by language strings indicating logical connections and the guidance of thought.

(26) Nor , for that matter, has the government . But even if industry provided funding, the inspections might falter or become misleading, which would be even worse for the public.

(27) from Brown text J33: Ralph B. Long, *The Sentence and Its Parts*

a. Dominant stress [in hymns] is of course more than extended duration, and normally centers on syllables that would have primary stress or phrase stress if the words or longer units they are parts of were spoken alone: a dominant stress given to "glorify" would normally center on its first syllable rather than its last . But the parallel is significant.

b. The personal pronouns and substitute "one" are normally unstressed because they refer to what is prominent in the immediate context. In "I'll go with George" dominant stress is probably on George ; but if George has just been mentioned prominently (and the trip to be made has been under discus-

sion), what is said is probably "I'll go with him," and dominant stress is probably on the preposition "with."

c. When a gesture accompanies "who's he?" the personal pronoun has dominant stress because "he" has not been mentioned previously.

d. If both George and a piece of information George does not have are prominent in the context , but the idea of telling George is new, then dominant stress will probably be on tell in why not tell George?

e. Thus the unstressed it of "it rarely snows here" gets its significance from its use with snows: nothing can snow snow but "it".

f. In "I have things to do" the word "things" makes little real contribution to meaning and has weaker stress than "do." If "work" is substituted for "things" (with more exact contribution to meaning), it will have dominant stress.

Share Social Ties

Writers use this category of language to prime shared knowledge of standards, resemblances, authorities, precedents, and values. The ancient rhetoricians had an umbrella name for this class of priors—commonplaces. Writers prime commonplaces in bids to increase a solidarity between themselves and their reader. Merely referencing a commonplace highlights the shared knowledge, beliefs and values that the writer might attempt to use

as leverage for persuasion. Such priming is often an efficient means of communication because commonplaces may function as implicit premises to carry conclusions (either positive or negative).

(28) It was decided our food distribution system will be safe only when security is assured. This should be done without resorting to obnoxious, maladapted policing practices that impinge farmers' liberty.

(29) from Brown text H11: U.S. Gov't, *1961 Research Highlights of the National Bureau of Standards*

a. Another recent achievement was the successful development of a method for the complete combustion in a bomb calorimeter of a metal in fluorine when the product is relatively non-volatile.

b. The accuracy of measuring the total electrical energy entering an exploding wire during a few microseconds was verified when two independent types of comparison with the heat energy produced had an uncertainty of less than 2 percent. This agreement is considered very good for such short time intervals.

c. This involves the application of a strong magnetic field to the radical vapor, which shifts the low-frequency spectra to a conveniently high microwave range, where they may be measured with optimum sensitivity.

d. These data are not of the precision obtainable by the methods previously mentioned, but the vast number of approximate

values available will be useful in many areas.

Direct Activity

Writers use these priming strings to address readers negotiating or contemplating actions in their proximate environments. The actions may be related to some institutional procedure or practice (like filling out a tax form), may require mental focus and review (like using a new computer program), or may direct physical actions with the reader's hands or body (like using the knarling attachment on a metal lave). Writers use this category of strings to help readers accomplish tasks by priming awareness of actions to be taken and confirmations of environmental changes that should procede and result directly from such actions.

(30) We must all be vigilent. Check for the organic purity of your food by rotating the package quickly and look for the new EU holographic symbol. If you cannot locate the symbol you can always return the package to your grocer.

(31) from Brown text E14: Ann Carnahan, *Nick Manero's Cook-out Barbeque Book*

a. Do start fires one or two hours ahead of time to obtain a lasting bed of glowing coals . Keep ashes from one barbecue to the next to sprinkle over coals if they are too hot, and to stop flames that arise from melting grease.

b. Don't forget to buy a plastic pastry brush for basting with sauces. Clean it meticulously in boiling water and detergent,

rinse thoroughly.

c. Also make sure thermometer does not touch the revolving spit or hit the coals.

d. Don't practice a new recipe on guests . Have a test-run on the family first, to be sure timing and seasoning are right.

e. Do buy meat the day or the day before you intend to cook it . Keep it no longer than 36 hours before cooking, and keep it in the coldest (but non-freezing) compartment of the refrigerator.

f. Don't think you have to start with the most expensive equipment in the world. The simplest grill (pan type) or inexpensive hibachi can make you a chef. You need tongs to handle meat; long forks for turning potatoes and corn; heavy foil on hand at all times. And lots of hot pads!

Interacting

Using these strings, writers acknowledge—in some cases fully address—the reader's presence, attempting to guide the reader's cognitive processing explicitly. Writers use these strings to prime reader understandings of the intentions and purposes behind the text and telegraphing a desire for reader responsiveness.

(32) Does this seem silly? Remember, the constitution shields us from such practices, for reasons I'll make clear to you.

(33) from Brown text E29: Edward Walton, "On Education for the Interior Designer"

a. It is true that most architectural schools have five year courses, some even have six or more. The element of public danger which enters so largely into architectural certification, however, would demand a prolonged study of structure. The real question that follows is—how are those four years used and what is their value as training ?

b. Without comparing the relative merits of the two courses—architecture versus interior design—let us examine the educational needs of the interior designer. To begin with, what is an interior designer? "The Dictionary of Occupational Titles" published by the U. S. Department of Labor describes him as follows...

c. The question, however, cannot be ignored for long. The basic problem involved is that a college setting up a graduate school must have an entirely separate faculty for the advanced degree. Most professors in the course must, naturally, again have a higher degree than the course offers. One solution is the aquisition of degrees in education but it is a poor substitute. It is a sort of academic ring-around-a rosy and you solve it.

d. This brings us to the question of accreditation of art schools in general.

Notifying

Writers use "notifying" strings to signal readers about the particularly salient information in the stream of information generated by a text. Writers implicitly indicate, for example, key terms and the presence of chunks of information that the writer believes will be important to the reader. Writers use this category to prime particular cognitive functionality for readers, helping them create schemata and implicitly conveying any conceptual organization within the information.

(34) The paradox is that the agency that does this is shielded by international courts with wide jurisdiction. By most definitions the agency would promise a bright future.

(35) from Brown text J31: Harold Searles, "Schizophrenic Communication"

a. It is not easy for the therapist to discern when, in the patient's communicating, an introject has appeared and is holding sway.

b. The content of his words may lapse back into monotonous repetition, as if a phonograph needle were stuck in one groove; only seldom is it so simple as to be a matter of his obviously parroting some timeworn axiom, common to our culture, which he has evidently heard, over and over, from a parent until he experiences it as part of him.

c. Eventually such incidents became more sporadic, and more sharply demarcated from her day-after-day behavior, and in

one particular session, after several minutes of such behavior—which, as usual, went on without any accompanying words from her—she asked, eagerly, "Did you see Granny"?

d. At first I did not know what she meant; I thought she must be seeing me as some one who had just come from seeing her grandmother, in their distant home-city. Then I realized that she had been deliberately showing me, this time, what Granny was like; and when I replied in this spirit, she corroborated my hunch.

e. At another phase in the therapy, when a pathogenic mother-introject began to emerge more and more upon the investigative scene, she muttered in a low but intense voice, to herself,"I hate that woman inside me"!

Linear Guidance

Writers sometimes guide readers rather explicitly through the linear text on the page. This is often done in an attempt to increase their readers' sense of familiarity by priming awareness of the path being followed through the text. Writers preview coming features of the text, provide connections back to previously-covered topics, and hint at how particular text should be perceived. Such guidance provides important grounding for readers within the text, helping them understand that the text was composed to make information familiar as needed for their progress through the text.

(36) Curiously, to avoid more of these problems the agency needs strong oversight to let it know its boundaries. By way of background, I will offer that, before the EU, each country insisted that her laws were preimenent.

(37) from Brown text P11: Bessie Breuer, *Take Care of My Roses*

a. SUCH a little thing to start with—the car registration. "Ida, where is the car license"? she asked. "I can't find it in the glove compartment".

b. " Via must have it ", I answered readily enough, recalling her last visit. " Via", she was frowning. "Why should Via have it"? Had she forgotten she had signed the car away, that whatever they mutually owned had been divided among the children? I was silent.

c. It is absurd of course to say that that one exclamation estranged me from the family I considered my very own, but there it hangs, a cooling void that broke our close connection with each other.

d. At the time I was filled with self-pity at this separation, but in the years since I have come to understand that the sight of me was painful to them after that outcry. In my person they would always remember that last long time of me alone with her, so if they told themselves that I could have prevented it, I can understand that by now and love them still, because everyone must justify, have a scapegoat for what is not to be borne.

e. The coffin stood on trestles in a corner of the long low dimly lit funeral parlor, on its dark shining surface the sheaf of white roses I had ordered. I knelt, just for decency I thought at the time, but found myself whispering, "Our Father which Art in Heaven..." And it was only after that that something unlocked in me and I felt a grief.

f. Again I felt impelled to kneel, and reached back and pulled Via down. Something would come into her heart... if nothing else the sounds of Bach would give her some healing.

4.5.3 Priming effects of the *description* cluster

The categories of this cluster prime a reader for visualizing physical presence. The presence might be a physical entitity, an occurrence, or some other phenomena. Such priming helps the reader perceive aspects of the world the writer is creating in the text.

Word Picture

Writers prime images that embody all the major elements in the story. These images allow readers to "see" the stuff of the story in mental imagery, providing a real sense of the entities (people, places, things, events) being represented by the text.

(38) The organization set about hiring 100 analysts in 56 European cities, looking for experienced inspectors to teach farmers the intricacies of food packaging.

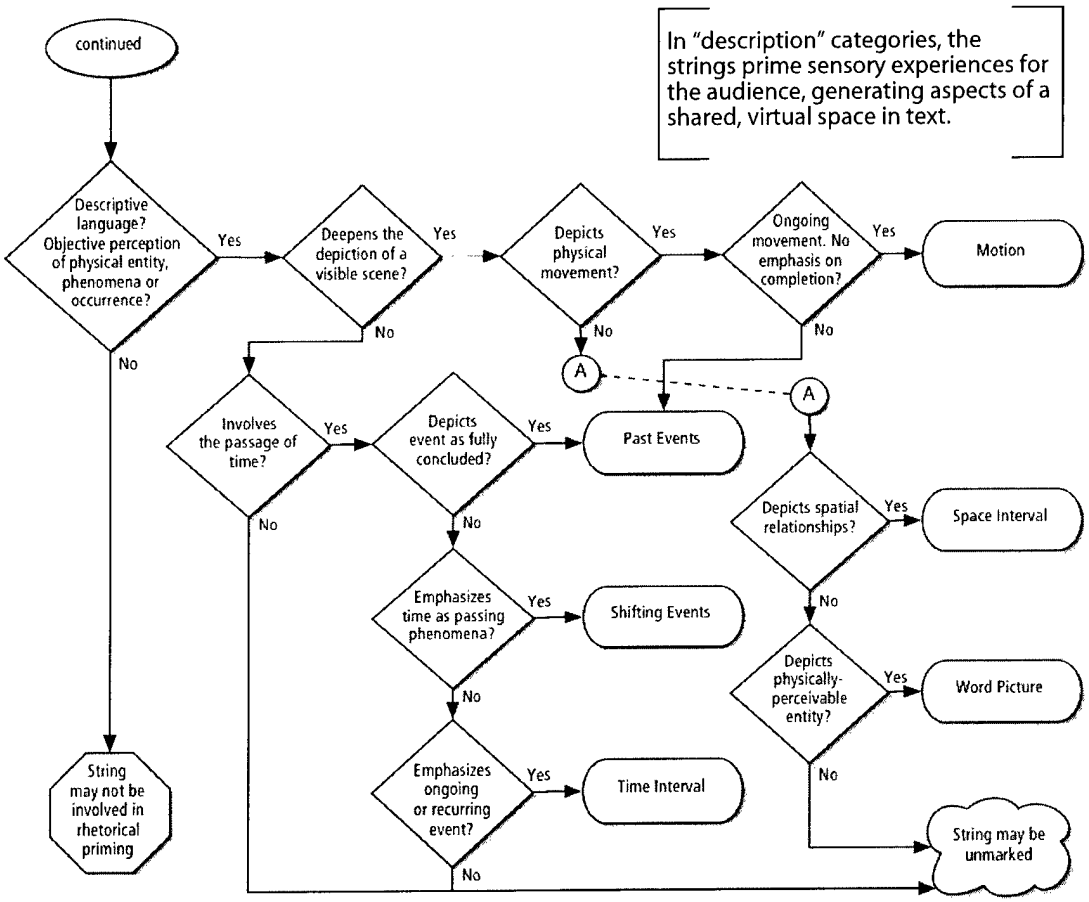


FIGURE 4.4: Heuristic diagram for categories of the *description* cluster

(39) from Brown text E14: Ann Carnahan, *Nick Manero's Cook-out Barbeque Book*

a. To keep the grill high enough above the fire so that when fat from meat drips down and flares up, flames cannot reach the meat.

b. Don't forget to have a supply of Melamine plates, bowls, cups, saucers, and platters for outdoor use. Made of the world's toughest unbreakable plastic, Melamine dinnerware comes in almost 400 different patterns and dozens of colors. There is even one set that has "barbecue" written on it.

c. Don't forget—when you take to the hills or the beach—that your cooler, which you might have used for wine- or beer-cooling on your terrace or back yard, is indispensable for carrying liquid refreshments.

d. There are many varieties of coolers and they serve many purposes. With them, you can carry steaks and hamburgers at refrigerator temperatures, and also get your frozen food for stews and chowders, to the marina or picnic, in A-1 condition. Do use paper napkins; lots of them.

Space Interval

Writers need spatial contiguity in their texts to help orient readers to the virtual world of the text. The strings in this category prime such spatial experiences, relying primarily on a handful of strings containing

prepositions that signal relationships between represented entities. Writers sometimes use this category to prime spatial relationships that are metaphorical in that they are relationships among abstract concepts.

(40) Along with the yearly compendium, a web site will accompany the paper edition. The two editions will be available next to one another. A new press office will be built near the Brandenburg Gate and will house a web kiosk with several chairs on its first floor.

(41) from Brown text N13: Martha Ferguson McKeown, *Mountains Ahead*

a. Over his shoulder he could see Max's loose grin and the Burnsides ' glowering faces. "Honey", he whispered. "Soon as we send them on their way and make camp, let's you and me go for a walk down by the Snake—all by ourselves".

b. Soon as the Burnsides moved on, he'd lead Rex down by the river; there he could shave and scrub himself up for the evening. Damn it, he thought bitterly, picking up his shirt and staring at the fresh bullet hole in the sleeve. If I hadn't got Nate stopped when I did, my duds'd all be shot plumb to hell!

c. He stooped, picked up his ruined hat, and pursed his lips thoughtfully. From the way the wound in his head was itching, Dan knew that it would heal. But his only hat was something else again." Nate! Nate"! he shouted. The Burnsides, now ready to roll, were purposefully deaf to his cry. "Nate"! he bellowed to the retreating back directly in front of him."I ain't going to fight you no more".

Motion

Writers use the strings of this category to prime kinetic images in the mind of the reader. These help readers build an image of ongoing motion. Writers specialize the kinetic shape of the images in their textual worlds, priming specific awareness of the movements taking place within.

(42) The reporters present scribbled notes and munched on organic goodies, scarfing down the food provided. They let out a collective gasp, choking and tittering as it dawned on them the director of France's *Renseignements Generaux* wanted to elbow his way into the enforcement regimen, rifling mail and tapping the phones of farmers and stalking those "dangerous" conspiritors who might duck the new law.

(43) from Brown text E16: Hal Kelly, "Build Hotei"

a. I used a Homemaster Routo-Jig made by Porter Cable for this job. It's good for cutting all the planking because it cuts with a bit-like blade at high rpm and does not chatter the plywood like a saber saw. When cut, the planking is clamped in place for a final and careful trimming.

b. Then lay a three-inch-wide strip of cloth along the keel line from the transom to the point of the stem. Before the resin has hardened, screw a one-inch mahogany keel strip along the centerline. This protects the bottom in beaching.

c. I laid three layers of glass cloth on the inside of the stem, also

installing a bow eye at this time.

d. After trimming off the excess on the frames and transom which was used to fasten them to the jig at a working height, the top of the side planking is installed.

Past Events

Writers must sometimes convey the elapse of time to readers. The simplest way is through the use of the simple past tense. This signals the reader that an event has started and stopped, without the necessity of placing the reader at the event. Thus, writers may collapse whole periods into closed-off events, efficiently conveyed.

(44) They caught a case of fraud immediately after the agency opened. It operated out of Germany, until the French saw the opportunity to garner more power and expelled the Germans from the council.

(45) from Brown text L05: Brett Halliday, *The Careless Corpse*

a. Then he turned the telephone over to Rourke, and went into the bedroom to change his slippers for dry socks and shoes. Rourke was talking on the phone when he came back.

b. "About an hour, eh? Are you positive"? He listened a moment and then said," Hold it".

c. Rourke hastily slopped whiskey into his glass on top of half-melted ice-cubes. " I'd better keep on driving yours", Shayne decided, " because I'll be going on over to the Beach.

d. I can drop you back here to pick mine up". He went to a closet to get a light jacket, and took his hat from beside the door. Timothy Rourke gulped down the whiskey hastily and joined him, asking," Who are we going to call on in the Northeast section"?

Time Interval

Writers use strings of this category to indicate event uniqueness or repetition over time. Writers prime awareness of repeated events through adverbial and prepositional strings, signaling recurrences and indicating temporal periods over which the textual events endure or recur.

(46) The agency is already changing. It got involved during the cold war when it was often used to squelch Soviet grain shipments every time they were headed to the middle east over the years.

(47) from Brown text J37: Douglas Ashford, "Elections in Morocco: Progress or Confusion?"

a. In analysis of the election falls naturally in four parts. First is the long and still somewhat obscure process of preparation, planning and discussion.

b. Preparation began slightly more than a year after independence with the first steps to organize rural communes. All political interests supported electoral planning, although there are some signs that the inherent uncertainties of a popular judgment led to some procrastination.

c. King Muhammad V was known to be most sympathetic to the formation of local self-government and made the first firm promise of elections on May Day, 1957. There followed a long and sometimes bitter discussion of the feasibility of elections for the fall of 1957, in which it appears that the Minister of the Interior took the most pessimistic view and that the Istiqlal was something less than enthusiastic.

d. Until the Charter of Liberties was issued in the fall of 1958, there were no guarantees of the right to assemble or to organize for political purposes.

Shifting Events

Writers can convey time's passage by creating shifts across discreet time horizons. Readers feel these shifts as a change in the specific scene or time period in which events of the text occurred (or will occur in the future). Writers use these strings to prime audience awareness of the passage of time.

(48) Once a food shipment enters the country, the food will be labeled immediately after it is inspected. A few hours later, the fruit will turn up at your local grocery. (At least that will occur when each inspector is fully trained this autumn) . Until then, be wary of "organic" labels.

(49) P10: Jay Williams, *The Forger*

a. I remember once I did a jacket for Magpie Press; the book was a fine historical novel about Edward III, and I did a week of research to get the details just right: the fifteenth-century armor, furnishings, clothes.

b. He also had, at times, an uncanny absent-minded air like a sleepwalker; he would look right through you while you were talking to him, and if you said, " For Christ's sake, Donald, you've got Prussian blue all over your shirt", he would smile, and nod, and an hour later the paint would be all over his pants as well.

c. He went into Mrs. Monmouth's library, which had low bookshelves all along the walls, and above them a Modigliani portrait, a Jackson Pollock twelve feet long, and a gorgeous Miro with a yellow background, that looked like an inscription from a Martian tomb.

d. " I don't think I've reached the point, yet , where I can say I know everything I ought to know about the craft. Besides, it's important to the way a painter thinks that he should move in a certain atmosphere, an atmosphere in which he may absorb the ideas of other masters, as Durer went to Italy to meet Bellini and Mantegna".

e. One of these days, I'm going to organize a gigantic exhibition that will span everything that's being painted these days, from extreme abstract expressionism to extreme photorealism, and

then you'll be able to see at a glance how much artists have in common with each other. I had studied with Burns ten years before, during the scholarship year the Manhattan gave me, along with the five-hundred-dollar prize for my paintings of bums on Hudson Street.

f. It was an awkward hour, but I didn't have to punch any time clock, and it only meant that sometimes I had to stay a couple of hours later at the drawing board to finish up a job.

g. After a short time, both George and Donald joined the class with me so they wouldn't feel lonely, and we used to hang a sign on the door of the Brush-off reading OUT TO WORK.

5 The Dimensions of Language Variation

This study's overarching purpose is to seek a set of language dimensions useful for understanding differences among the genres of writing in the 500-text Brown corpus (Francis & Kucera, 1979; Hofland et al., 1999). This chapter turns to a presentation and discussion of these language dimensions, suggesting some of the insights they afford on language usage. Each of the language dimensions is made up of several categories of text features defined by the theory of rhetorical priming (Kaufer et al., In Press).

The set of language dimensions presented in this chapter was initially derived by a statistical process known as "factor analysis," described in chapter 3. Five factors were extracted from data generated by an automated tagging of the corpus, excluding those texts in the "miscellaneous" genre. For convenience I've repeated the table showing the results of this process (table 5.1). It shows the rotated loadings for the five significant[1]

[1]Significant in the sense that each describes more variance than a single variable (i.e.

TABLE 5.1: The five significant factors

Factor 1		Factor 2	
Motion	0.816	Reasoning	0.771
Space Interval	0.790	Inner Thinking	0.753
Word Picture	0.715	Think Negative	0.692
Past Events	0.528	Linear Guidance	0.378
Linear Guidance	0.392	Share Soc Ties	0.368
Interacting	0.351		
(Share Soc Ties)	-0.597	(Time Interval)	-0.347
(Time Interval)	-0.591		
(Notifying)	-0.565		
Factor 3		**Factor 4**	
Think Back	0.752	(Think Ahead)	-0.802
Past Events	0.648	(Direct Activity)	-0.753
Linear Guidance	0.624	(Interacting)	-0.363
Shifting Events	0.589		
(Notifying)	-0.442		
(Direct Activity)	-0.337		
Factor 5			
(Think Positive)	-0.746		
(First Person)	-0.678		
(Interacting)	-0.415		
(Shifting Events)	-0.399		

factors.

The sections in this chapter each consider one of these factors, interpreting it as a language dimension and exploring its implications as seen in the Brown corpus genres, which are listed in table 5.2. Chapter 6 extends the discussion to consider how these language dimensions might

its eigenvalue > 1.0) and in the sense that each is affected by text genre at the $\alpha = 0.01$ level or better.

TABLE 5.2: The genres of the Brown and Frown corpora

Informative Prose Genres	Imaginative Prose Genres
A: Press Reportage	K: General Fiction
B: Press Editorials	L: Mystery and Detective Fiction
C: Press Reviews	M: Science Fiction
D: Religion	N: Adventure and Western
E: Skills and Hobbies	P: Romance and Love Story
F: Popular Lore	R: Humor
G: Belles Lettres, Biography, Memoir	
H: Miscellaneous	
J: Learned	

[1] Letters preceding each genre are used to refer to the genre
[2] The division of genres into "informative" and "imaginative" groups was made by the original compilers of the Brown corpus (see Francis & Kucera, 1979, p. 2)

be useful more generally, suggesting applications beyond these 15 Brown genres.

Each section in this chapter includes boxplots comparing the genres' factor scores to one another.[2] These visual representations are important because they help summarize the similarities and differences between the genres, crystallizing the authors' usage of each language dimension in the texts of each genre.

For example, look at figure 5.1(a) (p. 134). All the texts in each genre are represented by this graph. The reference letter for the genres are listed along the bottom of the graph, beneath the boxplots. Each boxplot includes all the scores of the texts in that particular genre.

You'll notice in figure 5.1(a) some of the boxplots are longer than oth-

[2]The data behind the boxplots are summarized in appendix E (p. 257).

TABLE 5.3: Five named language dimensions and the salient rhetorical priming categories comprising each

Writing for the Eyes vs. Informing
Motion
Space Interval
Word Picture
Past Events
Linear Guidance
(Share Soc Ties)
(Time Interval)
(Notifying)

Writing for the Intellect
Reasoning
Inner Thinking
Think Negative
Linear Guidance
Share Soc Ties

Retrospecting vs. Notifying
Think Back
Past Events
Linear Guidance
Shifting Events
(Notifying)

Instructing
(Think Ahead)
(Direct Activity)

Referencing Positive Relationships
(Think Positive)
(First Person)
(Interacting)
(Shifting Events)

ers. The length of the boxplot represents the variance of the texts in the genre. Texts making up the shorter plots (such as genre C, "press reviews") vary less than texts in the longer plots (such as genre A, "press reportage"). This means the authors of "press reviews" used the rhetorical priming features making up this language dimension more consistently than did the authors of "press reportage" texts. The shaded region of each boxplot represents the 95% confidence interval for the mean, based upon the pooled standard deviation for the genres.

Also notice some of the genres have median scores[3] above zero and others have medians below zero. For a bi-polar dimension, such as the one shown in figure 5.1(a), this indicates genres with medians above zero make heavier use of the positively-loaded rhetorical priming features while the genres with medians below zero make heavier use of the negatively-loaded priming features. This will become more clear in the sections that follow as I provide examples from the Brown corpus texts to help you understand each language dimension.

Table 5.3 shows the five named language dimensions, including the rhetorical priming features found to be salient to each. Notice that not all of the statistically significant features of the factors (table 5.1) were found to be salient to the language dimensions (table 5.3). I explain the reasoning for these decisions more fully in the following sections as I interpret each factor as a language dimension. All of the rhetorical priming features that were dropped from the factors had relatively lower factor

[3]The genre's median score is represented on the boxplot by the horizontal line at the center of the boxed region.

loading scores, indicating each was less important to the character of its factor. Furthermore, the priming features deemed not to be salient were found to function inconsistently across the texts of the corpus, indicating they could be artifactual results of the mathematical factor extraction process rather than indicative of the underlying language used by the authors. All of the rhetorical priming features with factor loading scores over 0.40 (in either the positive or negative direction) were found to be salient to the language dimensions.

The five language dimensions shown in table 5.3 are discussed in the following sections. Each discussion provides extended quotations from the Brown corpus texts to help you see the language dimension in action for yourself. The explanations rely on an understanding of the rhetorical priming categories, which were described in chapter 4.

5.1 Dimension 1: Writing for the Eyes vs. Informing

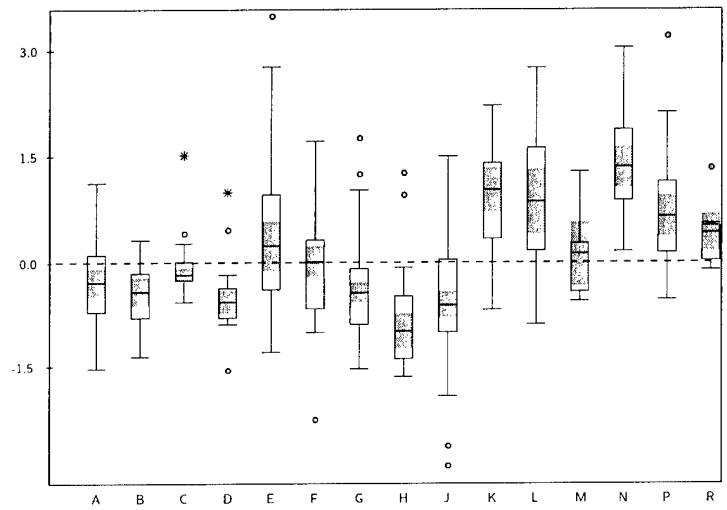
Consider the first factor in table 5.1. To interpret this factor as a dimension, we must consider how the co-occurring features (the positive and negative rhetorical priming categories) function within the texts to help achieve purpose. As you can see, four of the categories—"motion," "space interval," "word picture," and "past events"—have strong positive influence on the factor, all having loading scores above 0.50. The interpretation of this group of features is fairly straightforward, since all are descriptive in nature, helping the audience experience or "see" the virtual world the

author has created.

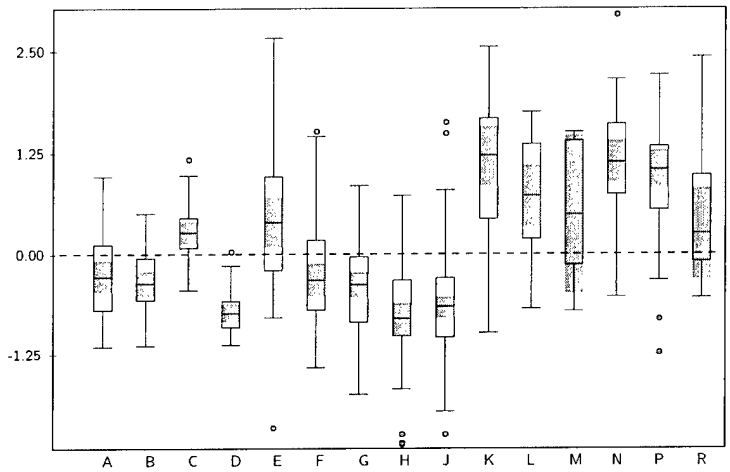
As you can see from the boxplots of this language dimension (figure 5.1), the authors of the "imaginative" texts (genres K–R at the right side of the figure) tended to make use of these positive categories, indicated by those genres' tendency toward higher positive scores. The high-scoring texts on this dimension seem to reduce the distance between the reader and the action of the text. This effect may be easily experienced in this passage from "Rattlesnake Ridge" by Peter Field (N14):

> He tramped out of the Miner's Rest with his hopes plummeting, and headed doggedly for the Palace Saloon, the last place of any consequence on this side of the street.
>
> The Palace was an elaborate establishment, built practically on stilts in front, with long flights of wooden steps running up to the porch. Behind its ornate facade the notorious dive clung like a bird's nest to the rocky ribs of the canyonside.
>
> Russ ran up the steps quickly to the plank porch. The front windows of the place were long and narrow, reaching nearly to the floor and affording an unusually good view of the interior. Heading for the batwings, Cobb glanced perfunctorily through the nearest window, and suddenly dodged aside. Nerves tight as a bowstring, he paused to gather his wits.

As we read this passage we visualize the described entities and see the movements and action of the character. Kaufer & Butler (2000, chapter 3) describe this as "the feel of extended space" (p. 83), describing several



(a) Brown Corpus texts (pooled $SD = 0.73$)



(b) Frown Corpus texts (pooled $SD = 0.68$)

FIGURE 5.1: Boxplots showing dimension 1, "Writing for the Eyes vs. Informing," scores for the text genres

of the techniques authors use to create "immediacy" (p. 85)—the feeling that the reader is within the world of the text. This is a feeling anyone who has ever been scared under the covers by a frightening passage in a suspense novel can identify (cf. Olson, 1994).

Maintaining such a world in the text and ensuring readers' engagement with it requires fluid movement through the text. A reader who is continually getting lost in a text would not feel immediacy and the illusion of inhabiting the text's virtual world would be broken. One of the ways writers attempt to ensure this doesn't happen is through their use of "linear guidance" priming. This priming category is also associated positively with this first language dimension, although not as strongly as the others (with a loading score of 0.39). This priming category is found to some degree in most examples of highly-visual text in the corpus, working to maintain reader familiarity and connections within the text.

For example, notice the use of connecting pronouns and contiguous descriptions that lend a feeling of unified visible and aural space for a reader of this passage from "Aid and Comfort to the Enemy" by Wesley Newton (N25):

> EARLY that day Matsuo saw a marine. The enemy came looming around a bend in the trail and Matsuo took a hasty shot, then fled without knowing the result, ran until breath was a pain in his chest and his legs were rubbery. As his feet slowed, he felt ashamed of the panic and resolved to make a stand. He crossed the next meadow and climbed a tree where the jun-

> gle trail resumed. In the leafiest part of the tree, straddling a broad horizontal limb, he could see over the meadow. . . .
>
> Fresh on his mind were events of the past day when his whole regiment was destroyed in the hills. They had fought from caves, and the marines resorted to burning them out. Even now, like a ringing in his ears, he heard the wooooosh of flame-throwers squirting great orange billows. A wave of flame rippling through their cave had reached Nagamo, his friend, and with a shriek the man bolted through the entrance, then slowed to the jerky walk of a puppet, his uniform blazing.

We see this use of "linear guidance" priming further exemplified in a passage from Edwin Booth's *Outlaw Town* (N12) in which the author must let the reader know to which character to attribute the speech and action taking place within the text:

> "One thing, Summers", Brenner said. "You're not to mention my name. Tell her Curt Adams wants to see her".
>
> Summers pulled up short, and turned around. "I don't know, Mr. Brenner", he said haltingly, beginning to get an inkling of Brenner's plans. "It doesn't seem quite right, telling her a thing like that. Couldn't I just"—His voice trailed off into silence. Brenner continued to smile, but his eyes were cold.

In addition to the "imaginative" genres, we also find authors using the positive features of this language dimension in some of the texts of the "informative" genres (A–J). One of these genres, "skills and hobbies,"

has a tendency toward the descriptive (i.e. its mean is above zero on this dimension). Many of the texts of this genre, like those of the fiction genres, seem written to help a reader experience contiguous space, although in the case of "skills and hobbies," the usage often helps the reader visualize navigation in the space of the text, comparing it to what the reader may experience outside the text.

For example, in an article from *McCall's Needlework and Crafts* (E15), the author describes the tools, materials, and procedures for creating clay tiles, carefully directing readers in the motions they're to follow through the process in their own environment:

> MATERIALS: Ceramic modeling clay: red, white or buff. Stoneware clay for tiles. Glazes, one-stroke ceramic colors, stains, cones as indicated in the individual instructions.
>
> GENERAL DIRECTIONS: Use well-wedged clay, free of air bubbles and pliable enough to bend without cracking. Clean wooden molds and presses thoroughly; they must be free of oil, wax and dust.
>
> PRESSING DESIGNS: The size of wooden mold will determine the amount of clay needed. Roll clay to thickness indicated in individual instructions. Whenever possible, use the wooden mold as a pattern for cutting clay. When mold has more than one design cavity, make individual paper patterns. Place mold or paper pattern on rolled clay and cut clay by holding knife in vertical position (cut more pieces than required for

> project to make allowance for defects; experiment with defects for decoration techniques of glazes and colors).

As readers of this passage, we are told to use a clean wooden mold, but the author doesn't leave it at that, the author goes on to describe what a clean mold looks like, "free of oil, wax and dust."

The fifth positive rhetorical priming category in this factor, "interacting," is somewhat weaker than the others (with a loading score of 0.35) and is used differently than the other categories of this language dimension, suggesting it may not be salient to the dimension.

For example, later in Newton's text about marines in World War II (N25), his narrator indicates a tentative thought with the phrase "A signal?", providing a touch of interactivity (indicated by the question mark). This reveals the thought of the character, but it also suggests the thought could be one the reader should experience in a way that simply relaying the thought (e.g., "A signal, he thought.") would not:

> As [Matsuo] looked up from picking at a leg ulcer, he saw a marine in the jungle across the clearing. Gloom receded. The marine came to the edge of the green jungle mist and stayed, as though debating whether to brave the sunlight. His fatigues made a streak of almost phosphorescent green in the mist. "Come out, come out in the meadow", Matsuo said under his breath. The man leaned against a tree and wiped a sleeve across his face. A signal? Matsuo lifted his rifle, easing the sling under his left upper arm for steadiness.

We find other indications of "interacting" priming in imaginative texts that are heavy with dialog. Characters engaged in dialog refer to one another and ask questions of other characters. Such instances do not attempt to interact with the reader of the text, but may indicate relationships among characters in the text. In this regard, this usage of "interacting" priming doesn't seem consistent with the other priming categories of this language dimension.

This language dimension is bi-polar, meaning the factor on which it is based has both positive and negative priming features. Considering the co-occurring, negatively-loaded features in table 5.1, we find three categories—"share social ties," "time interval," and "notifying"—that have strong influence on the factor (the magnitudes of their loading scores is above 0.50).

Exploring the use of these rhetorical priming categories in the Brown corpus, it seems writers combine them to maintain or even attempt to magnify a separation between the reader and the entities in the writing. Writers making use of the negative priming categories seem to be informing the reader about the subject, rather than asking the reader to experience the subject first hand.

For example, in an academic essay by Brand Blanshard (J52), the author attempts to convince the reader of his point by evoking "social ties"—a priming category that is characteristic of many non-fiction forms of argumentation—and by using "notifying" priming strings to help the reader follow along and organize the information being presented:

> The [emotive] theory claims to show by analysis that when we say, "That is good", we do not mean to assert a character of the subject of which we are thinking. I shall argue that we do mean to do just that.
>
> Let us work through an example, and the simpler and commoner the better. There is perhaps no value statement on which people would more universally agree than the statement that intense pain is bad. Let us take a set of circumstances in which I happen to be interested on the legislative side and in which I think every one of us might naturally make such a statement.

Such language invoking social ties with the reader are often found in the "informative" texts, helping readers see what the author means and agree without the need for creating immediacy. Instead, such texts maintain what Walter Chafe (1994) calls "displacement," keeping the reader removed from a world described in the text and, instead, informing them of the conclusions to be drawn from it.

Displacement may be further heightened by "time interval" priming—informing readers of time periods, as opposed to letting these be experienced through "shifting events" priming—and using "notifying" cues, which, as we saw in the above example, tend to increase awareness of information flow, rather than emersion in a scene. Of course emersion is further controlled by providing little in the way of description, the key positive priming categories of the language dimension.

For example, Clark Vincent's text on "unmarried mothers" (F39) maintains a distance between reader and subject throughout. In the following quotation Vincent delivers his conclusion in a way that a reader of fiction would probably find disconcerting for its lack of description and its use of the "time intervals" and "notifying" categories:

> This conclusion is dependent on the assumption that traditional sex mores will continue to sanction both premarital chastity as the "ideal", and the double standard holding females primarily responsible for preserving the ideal.
>
> Our discussion of this involves using Erik Erikson's schema of "identity vs. identity diffusion" as a conceptual tool in superimposing a few common denominators onto the diverse personality and family configurations of the unwed mothers from whose case histories we quoted earlier. . . . We hope thereby to emphasize that, from a psychological standpoint, the effectual prevention of illegitimacy is a continuous long-term process involving the socialization of the female from infancy through adolescence.
>
> Hypothesizing a series of developmental stages that begin in the individual's infancy and end in his old age, Erikson has indicated that the adolescent is faced with a series of identity crises. The successful and positive resolution of these crises during adolescence involves an epigenetic principle—during adolescence, the individual's positive resolutions in each area

> of identity crisis depend, to a considerable degree, on his already having resolved preliminary and preparatory identity crises during his infancy, childhood, and early adolescence.

Texts in the corpus seem especially likely to make use of these negatively-loaded priming categories as they summarize or draw broad generalizations about ideas and concepts. For example, Newton Stallknecht and Frank Oppenheimer use all three of the negative priming categories in their texts regarding academic disciplines and progress. Stallknecht (G44) writes:

> History has this in common with every other science: that the historian is not allowed to claim any single piece of knowledge, except where he can justify his claim by exhibiting to himself in the first place, and secondly to any one else who is both able and willing to follow his demonstration, the grounds upon which it is based. This is what was meant, above, by describing history as inferential. The knowledge in virtue of which a man is an historian is a knowledge of what the evidence at his disposal proves about certain events.
>
> It is obvious that the historian who seeks to recapture the ideas that have motivated human behavior throughout a given period will find the art and literature of that age one of his central and major concerns, by no means a mere supplement or adjunct of significant historical research. The student of ideas and their place in history will always be concerned with

> the patterns of transition, which are at the same time patterns of transformation, whereby ideas pass from one area of activity to another.

And Oppenheimer (G11), writing on the need for "scientific progress" in the "soft" sciences, writes:

> Social invention did not have to await social theory any more than use of the warmth of a fire had to await Lavoisier or the buoyant protection of a boat the formulations of Archimedes. But it has been during the last two centuries, during the scientific revolution, that our independence from the physical environment has made the most rapid strides.
>
> We have ample light when the sun sets; the temperature of our homes is independent of the seasons; we fly through the air, although gravity pulls us down; the range of our voice ignores distance. At what stage are social sciences then? Is the future of psychology akin to the rich future of physics at the time of Newton? There is a haunting resemblance between the notion of cause in Copernicus and in Freud. And it is certainly no slight to either of them to compare both their achievements and their impact....
>
> Just as present technology had to await the explanations of physics, so one might expect that social invention will follow growing sociological understanding. We are desperately in the need of such invention, for man is still very much at the mercy

> of man. In fact the accumulation of the hardware of destruction is day by day increasing our fear of each other.

Oppenheimer doesn't describe the "hardware of destruction" for his readers. Nor does he describe or let his readers experience the hardware's destructive effects that might lead to deductive understandings of the "increasing fear" he mentions in his last sentence. Describing each of these would require considerable space in the writing. Instead, Oppenheimer chooses to inform his readers about these things, assumes they get it, and then moves on, as do the authors of other texts making heavy use of the three negatively-loaded priming categories of this language dimension.

Interestingly, the authors of the "press reviews" genre seem to balance their usage of the positive and negative priming categories. For example, in the following review from the *Chicago Daily Tribune* (C01), we find an author using vivid prose to describe a performance. But the author puts this description into perspective for readers by informing them of comparisons with other performances and with expectations:

> IT IS NOT NEWS that Nathan Milstein is a wizard of the violin. Certainly not in Orchestra hall where he has played countless recitals, and where Thursday night he celebrated his 20th season with the Chicago Symphony orchestra, playing the Brahms Concerto with his own slashing, demon-ridden cadenza melting into the high, pale, pure and lovely song with which a violinist unlocks the heart of the music, or forever finds it closed.

> There was about that song something incandescent, for this Brahms was Milstein at white heat. Not the noblest performance we have heard him play, or the most spacious, or even the most eloquent. Those would be reserved for the orchestra's great nights when the soloist can surpass himself.
>
> This time the orchestra gave him some superb support fired by response to his own high mood. But he had in Walter Hendl a willing conductor able only up to a point. That is, when Mr. Milstein thrust straight to the core of the music, sparks flying, bow shredding, violin singing, glittering and sometimes spitting, Mr. Hendl could go along. But Mr. Hendl does not go straight to any point. He flounders and lets music sprawl. There was in the Brahms none of the mysterious and marvelous alchemy by which a great conductor can bring soloist, orchestra and music to ultimate fusion. So we had some dazzling and memorable Milstein, but not great Brahms.

We find an equally interesting contrast within the "learned" genre of texts. Although many of the texts tend to make use of the negative priming features of this language dimension, some—especially those from the physical sciences—have highly descriptive passages. For example, the author of the following quotation describes a heat transfer experiment (J02), helping the audience visualize the occurrences, possibly so the experiment can be replicated or perhaps to build trust in the author's characterization of the experimental results:

> The argon flow through the porous anode was varied systematically between fmula[4] and fmula. The lower limit was determined by the fact that for smaller flow rates the arc started to strike to the anode holder instead of to the porous graphite plug and that it became highly unstable. The upper limit was determined by the difficulty of measuring the characteristic anode surface temperature (see below) since only a small region of the anode was struck by the arc. This region which had a higher temperature than the rest of the anode surface changed size and location continuously.
>
> For each mass flow rate the arc voltage was measured. To measure the surface temperature of the anode plug, the surface was scanned with a pyrometer. As it turned out, a very hot region occurred on the plug.

Such descriptive passages from some academic texts may be contrasted with informative passages that are more typical of the "learned" texts in the corpus. For example, the following passage describes the findings of a psychological experiment (J28), informing the reader of the hypothesis and findings:

> It was predicted that Kohnstamm-negative subjects would adhere to more liberal, concretistic reports of what the ambiguous figure "looked like" as reflecting their hesitancy about tak-

[4]As described in appendix B, the string "fmula" was substituted for all formulæ appearing in the original corpora texts to prevent the software from incorrectly tagging portions of a formula as part of priming strings.

> ing chances.... Individual differences in Kohnstamm reactivity to controlled Kohnstamm situations were found among the subjects used in the study. Only 27 % (11 subjects) gave a positive Kohnstamm reaction when completely naive concerning the phenomenon.

We find similar usage of the negative priming categories in the following passage from an economics text by James O'Leary (J40) analyzing the interest rate outlook. Notice his usage of the "notifying" priming—used to signal the reader about information processing—and his use of "social ties" priming, priming to convey socially-shared assumptions without needing to describe all the underlying conditions and values explicitly:

> IN assessing the outlook for interest rates in 1961, the question, as always, is the prospect for general business activity. By and large, what happens to business as a whole will govern the relationship between demand and supply conditions in the capital markets and will thus determine interest rates. Moreover, the trend of general business activity in 1961 will exert a decisive influence on fiscal, monetary, and other Federal policies which affect interest rates. Nineteen-sixty has been a baffling year for analysts of general business activity. During much of the year the general level of business activity has moved along on a record-high plateau, but there have been persistent signs of slack in the economy. The tendency for general business activity to soften somewhat is becoming more evident.

After careful consideration of the interactions between the rhetorical priming categories involved in this language dimension, I named it "Writing for the Eyes vs. Informing." This name reflects the dimension's bipolar nature: there are two sides to the dimension. The first, positive set of categories follows Plato, Aristotle and many others who have taught the power of language intended to put images on the hearers' retinas: "writing for the eyes."

The other side, comprised of the negative priming categories, follows Richard Lanham, Joseph Williams and others who have noticed (and often recommended against) the way authors of information-heavy texts distance their readers (and themselves) from the action and entities of the text: "informing."

This is a good place for me to remind the reader this study is intended to be descriptive, not prescriptive: the findings presented here describe what the authors of the Brown and Frown corpora texts did, not what they should or should not have done. The "positive" and "negative" labels used in this study are not evaluative, but indicate the sign of the factor loading score for rhetorical priming categories within each language dimension.

Remember, the analyzed texts are all from published works and, it is assumed, the authors who wrote them needed to achieve some goal with their texts and intended them to be the way they are. The question this study set out to answer was whether or not there were detectable, interpretable differences in the texts' usage of micro-level rhetorical priming that was related to text genre. The answer to this question—demonstrated by each of the five sections of this chapter—is "yes." What

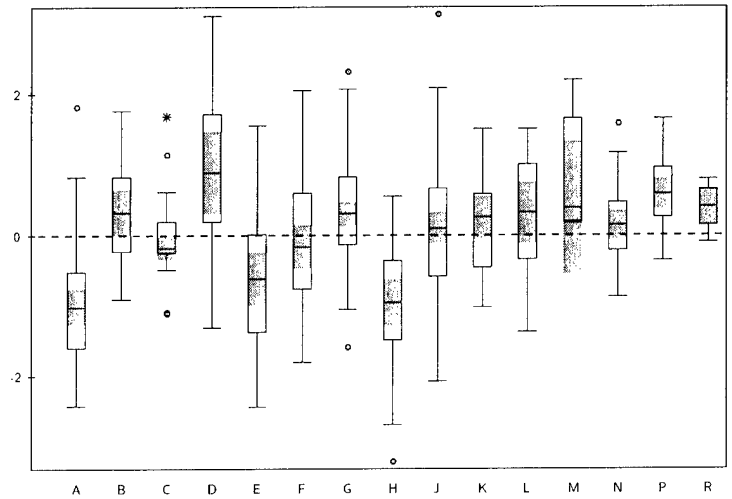
we are to do about it is a topic taken up in chapter 6.

5.2 Dimension 2: Writing for the Intellect

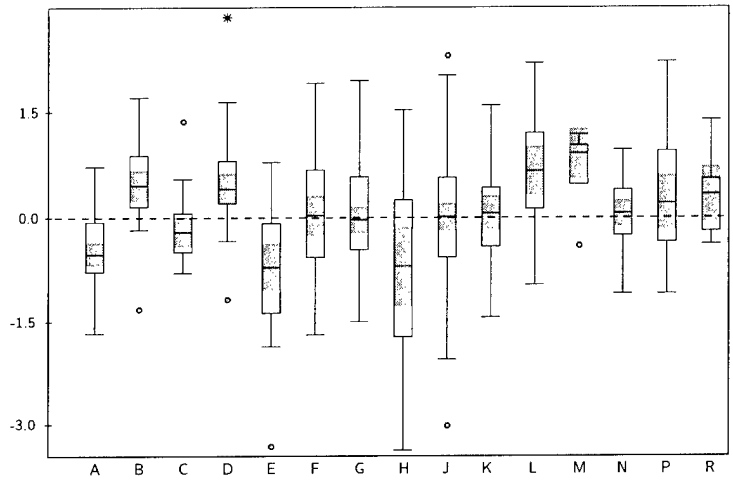
Consider the second factor in table 5.1 (p. 128). The salient priming categories on the second language dimension are "reasoning," "inner thinking," "think negative," "linear guidance," and "share social ties," all with positive loading scores. The first three have especially strong influence on the dimension, each with a factor loading score over 0.65. The co-occurrence of these three highly-marked categories is found in passages of the corpus that tend to share thinking directly with the reader, sometimes in overtly persuasive passages (found mostly in the "informative" genres) and sometimes in passages that simply present a line of reasoning for reader inspection and understanding or empathy.

Two passages from a travel article by Robert Deardorff (E13) provide instructive contrasts in the usage of this language dimension. This first quotation below makes use of the priming categories of the dimension, sharing Deardorff's thinking with the reader and leading the reader's mind through the perceptions the author shares:

> BUILT UPON seven hills, Istanbul, like Rome, is one of the most ancient cities in the world, filled with splendor and contrast. It is an exotic place, so different from the ordinary that the casual tourist is likely to see at first only the contrast and the ugliness of narrow streets lined with haphazard houses.



(a) Brown Corpus texts (pooled $SD = 0.84$)



(b) Frown Corpus texts (pooled $SD = 0.81$)

FIGURE 5.2: Boxplots showing dimension 2, "Writing for the Intellect," scores for the text genres

> At the moment, many of these are being pulled down. Whole blocks are disappearing and more are scheduled to vanish to make room for wide boulevards that will show off its treasures to better advantage—the great domes and graceful spires of its mosques, the panorama of the Bosphorus and the Golden Horn. Even when they are finished, however, the contrast will remain, for Istanbul is the only city in the world that is built upon two continents. For almost 3, 000 years Europe and Asia have rubbed shoulders in its streets.

In this passage Deardorff does not just inform his reader of his thinking, but he primes the reader to follow his reasoning. That is, we're not just informed that Istanbul is ancient, but that it is one of the most ancient cities, "like Rome." We're not just informed it is an exotic place, but we're led to understand it is "so different from the ordinary" that what we see at first may mislead us. Deardorff attempts to reason with us in this passage about the experience his text conveys.

In contrast, consider this second passage that immediately follows the above quotation. Deardorff almost seems to switch off his use of the second language dimension and, instead, he moves to an historical overview, relaying the happenings while providing no "reasoning" priming:

> Founded in the Ninth Century B.C. it was called Byzantium 200 years later when Byzas, ruler of the Megarians, expanded the settlement and named it after himself. About a thousand years after that, when the Roman Empire was divided, it be-

> came capital of the Eastern section. On May 11, 330 A.D., its name was changed again, this time to Constantinople after its emperor, Constantine. In 1453 when the last vestige of ancient Roman power fell to the Turks...

We don't know what Deardorff thinks about this history he's leading us through because he doesn't reveal his thinking about it to us. We probably trust him that the information is relevant to the entire text, but as readers we're not given any explicit indications to guide our understanding of this. Unlike the first quotation, Deardorff doesn't use "inner thinking" or "think negative" priming cues and he doesn't make any evident effort to prime "reasoning." In this passage he simply recounts history.

One commonality of the above passages, is their usage of the fourth salient category of priming associated with this dimension, "linear guidance." Like the writers we saw in factor 1 attempting to immerse the reader in scenic description, writers attempting to bring reader thinking along on an intellectual journey cannot afford to "lose" the reader. In the Deardorff quotations we see how adding touches of "linear guidance" seems to help the reader process the text. Perhaps this helps readers unveil any reasoning being delivered by the author.

For example, notice the usage of "linear guidance" priming in the following passage, helping to move the reader through the inter-connected sentences and joining together bits of reasoning. This passage by James J. Maguire (A35) leaves little to chance, engaging the reader's information processing at each step and helping enable the cogency of argument (even

if one ultimately disagrees with his reasoning):

> THERE ARE, so my biologist friends tell me, mechanisms of adaptation and defense that are just too complete and too satisfactory. Mollusks are a case in point. The shell, which served the strain so well at a relatively early stage in the evolutionary scheme, tended to cancel out the possibility of future development. Though this may or may not be good biology, it does aptly illustrate the strength and the weakness of American Catholic higher education.
>
> There can be no doubt that the American Catholic accomplishment in the field of higher education is most impressive: our European brethren never cease to marvel at the number and the size of our colleges and universities. The deeper wonder is how this miracle was accomplished in decades, rather than in centuries and by immigrant minorities at that. By way of explanation we ourselves are prone to imagine that this achievement stems from the same American Catholic zeal and generosity which brought the parochial school system into existence. There is, however, one curious discrepancy in this broad and flattering picture. Viewing the American Catholic educational achievement in retrospect, we may indeed see it as a unified whole extending from grade school to university.
>
> But the simple truth is that higher education has never really been an official American Catholic project; certainly not

> in the same sense that the establishment of a parochial school system has been a matter of official policy.

This passage, from a newspaper article, is unusual for its genre, using more of this language dimension than most "newspaper reportage" texts. Most texts in this genre limit the amount of thinking revealed, perhaps adhering to the old journalist saw, "we report, you decide." Whatever the reason, most of the texts in the "newspaper reportage" genre do not use this language dimension as frequently as the Maguire text quoted above does. His text's score on this is more typical of the texts of the "religion" genre.

As an example of that genre, notice the thinking and reasoning cues in this passage as the author, Peter Eldersveld (D07), marshals anecdotal and expert evidence to suggest a technique for helping alleviate parishioners' fear of fear itself:

> A certain teacher scheduled a "Fear Party" for her fourth grade pupils. It was a session at which all the youngsters were told to express their fears, to get them out in the open where they could talk about them freely. The teacher thought it was so successful that she asks: "Wouldn't it be helpful to all age groups if they could participate in a similar confessional of their fears and worries"?
>
> Dr. George W. Crane, a medical columnist, thinks it would. He says: "That would reduce neurotic ailments tremendously. . . . One of the most wholesome things you could schedule in your

> church would thus be a group confessional where people could admit of their inner tensions". We are evidently trying hard to think of new ways to deal with the problem of fear these days.
>
> It must be getting more serious. People are giving their doctors a hard time. One doctor made a careful survey of his patients and the reasons for their troubles, and he reported that 40% of them worried about things that never happened; 30% of them worried about past happenings which were completely beyond their control; 12% of them worried about their health, although their ailments were imaginary; 10% of them worried about their friends, neighbors, and relatives, most of whom were quite capable of taking care of themselves. Only 8% of the worries had behind them real causes which demanded attention. Well, most of our fears may be unfounded, but after you discover that fact, you have something else to worry about: Why then do we have these fears?

For an even more evident example, consider the following passage from Paul Ramsey's book, *War and the Christian Conscience* (D11). Ramsey's makes use of the language dimension as he attempts to help his reader navigate a theological justification of the doctrine of mutually-assured destruction:

> When they say that under no circumstances would it ever be right to "permit" the termination of the human race by human action, because there could not possibly be any propor-

> tionate grave reason to justify such a thing, they know exactly what they mean. Of course, in prudential calculation, in balancing the good directly intended and done against the evil unintended and indirectly done, no greater precision can be forthcoming than the subject allows.
>
> Yet it seems clear that there can be no good sufficiently great, or evil repelled sufficiently grave, to warrant the destruction of mankind by man's own action. I mean, however, that the moral theologian knows what he means by "permit". He is not talking in the main about probabilities, risks and danger in general. He is talking about an action which just as efficaciously does an evil thing (and is known certainly and unavoidably to lead to this evil result) as it efficaciously does some good. He is talking about double effects, of which the specific action causes directly the one and indirectly the other, but causes both; of which one is deliberately willed or intended and the other not intended or not directly intended, but still both are done, while the evil effect is, with equal consciousness on the part of the agent, foreknown to be among the consequences. This is what, in a technical sense, to "only permit" an evil result means.

As these last two examples amply demonstrate, much of the thinking revealed in the use of this language dimension is primed by the "think negative" category. That is, the micro-rhetorical choices seem to prime

awareness of thoughts many people would find undesirable. The co-occurrences of "reasoning" with "inner thinking" and "think negative" (but not "think positive") in the Brown corpus may be an artifact of the times (the cold war).

A tentative answer comes by examining the usage of this language dimension in the Frown corpus and comparing it to the usage in the Brown corpus. It appears the association between the priming categories may reflect the times in which the texts were composed, at least for the texts placed in the "religion" genre. The co-occurrences of the priming categories is still present in the 1990s texts, but the authors of texts in the "religion" genre, in particular, have reduced their usage of the language dimension.

Perhaps the easing of cold war tensions between the corpora led the "religious" texts away from their negative focus on nuclear holocaust. By the time the texts of the Frown corpus were composed, the "religious" genre was more similar to the "newspaper editorials," another genre in which the typical texts of the genre make use of this language dimension.

For example, in a short editorial letter taken from *The Philadelphia Inquirer* (B19), we find the writer using the rhetorical priming categories of this language dimension, indicative of attempts at shared thinking. This text also exemplifies the usage of "share social ties" priming that is associated with this language dimension. This priming can substitute for longer chains of reasoning, but these carry built-in warrants with premises that are socially shared:

> There is a trend today to bemoan the fact that Americans are too "soft". Unfortunately, those who would remedy our "softness" seek to do so with calisthenics. They are working on the wrong part of our anatomy. It is not our bodies but our hearts and heads that have grown too soft.
>
> Ashamed of our wealth and power, afraid of so-called world opinion and addicted to peace, we have allowed our soft-heartedness to lead to soft-headed policies. When we become firm enough to stand for those ideals which we know to be right, when we become hard enough to refuse to aid nations which do not permit self-determination, when we become strong enough to resist any more drifts towards socialism in our own Nation, when we recognize that our enemy is Communism not war, and when we realize that concessions to Communists do not insure peace or freedom, then, and only then will we no longer be "soft". America doesn't need to "push-up", she needs to stand up!

And we find a similar usage of the language dimension in this short piece, also from the same text sample from *The Philadelphia Inquirer* (B19):

> I wish to advocate two drastic changes in Washington Square:
>
> 1. Take away George Washington's statue.
>
> 2. Replace it with the statue of one or another of the world's famous dictators.

> There's no sense in being reminded of times that were. Washington Square seems not part of a free land. It may remind one of Russia, China or East Berlin; but it can't remind one of the freedom that Washington and the Continental soldiers fought for. The Fairmount Park Commission will no doubt approve my two proposals, because it is responsible for the change of ideological atmosphere in the Square.

It's worth noting that the satire of this piece is pretty obvious to a reader: the writer does not seriously suggest Washington's statue be replaced by a dictator's (the editorial writer goes on to complain that birds have been shoed from the park by the park commission, a "start on the road to Totalitarianism," according to the writer). Even though this dripping irony is somehow clear to the human reader, the usage of rhetorical priming categories in the two quotations above is nearly identical. That is, the serious and the ironic editorial use the same priming categories in leading the reader along the path of reasoning (cf. Kiley & Shuttleworth, 1971; Kaufer, 1981).

Heretofore in this section the discussion has focussed on the nonfiction genres. This is not to suggest the "imaginative" genres do not make use of language dimension 2. The authors of fiction in the Brown corpus also lead their readers through reasoned thoughts and judgments. This is not surprising to those familiar with Wayne Booth's (1961) *Rhetoric of Fiction.* Booth explores the extent to which authors insert themselves into the descriptive world of the novel in order to sway audience opinion.

For example, the texts of the "science fiction" genre are among those that seem to present characters' reasoning carefully, sometimes so the audience sees the broader implications of decisions, but sometimes so the audience might follow along with the plot. After all, following a plot involves understanding and ostensibly believing the characters' motivations. As you might expect, fiction authors don't leave this to chance. Instead, they combine "reasoning" priming for the audience with revelations of thoughts to lead their readers' thinking in a designed manner. This is evident in passages from James Blish's *The Star Dwellers* (M03):

> What Gabriel was being asked to do now, however, was to re-examine all his basic assumptions, make value-judgments on them, and give them new and different powers in his mind to govern his motives. This is not wholly a reasoning process—a computer cannot do it all—and even in an Angel it takes time. (Or, perhaps, especially in an Angel, whose assumptions had mostly been fixed millions of years ago.)
>
> Being reasonably sure of the reason for the long pause, however, did not make it seem any less long to Jack. He had already become used to Hesperus' snapping back answers to questions almost before Jack could get them asked. There was nothing he could do but wait. The dice were cast. At last Gabriel spoke. "We misjudged you", he said slowly.

And, following Gabriel's pronouncement, we are again asked to follow a chain of reasoning as the main character, a law clerk named Jack, decides

what to do:

> What Gabriel was asking was that mankind forego all its parochial moral judgments, and contract to let the Angels serve on Earth as it is in Heaven regardless of the applicable Earth laws. The Angels in turn would exercise similar restraints in respect for the natural preferences and natures of the Earthmen—but they had no faintest notion of man's perverse habit of passing and enforcing laws which were contrary to his own preferences and violations of his nature.
>
> The simple treaty principle that Gabriel was asking him to ratify, in short, was nothing less than total trust. Nothing less would serve. And it might be, considering the uncomfortable custom the Angels had of thinking of everything in terms of absolutes, that the proposal of anything less might well amount instead to something like a declaration of war.
>
> Furthermore, even the highly trained law clerk who was a part of Jack's total make-up could not understand how the principle could ever be codified. Almost the whole experience of mankind pointed toward suspicion, not trust, as the safest and sanest attitude toward all outsiders. Yet there was some precedent for it....

To be sure not all (or even most) of the "imaginative" texts in the corpus interspace such long passages of reasoning into their largely descriptive worlds. Nonetheless, many texts in the fiction genres make use of the

categories of this language dimension, as can be seen in figure 5.2(a).

For example, in this short passage from Jim Thompson's *The Transgressors* (N09) we find an omniscient narrator narrating the thoughts of one of the characters interweaved with the action and dialog of the text:

> Beaming idiotically, he pooched out his lips and attempted to kiss her. She yanked away from him furiously. "You shut up! shu-tt up-pp! I've got something to say to you, and by God you're going to listen. Do you hear me? You're going to listen"!
>
> Lord nodded agreeably. He said he wanted very much to listen. He knew that anything a brainy little lady like her had to say would be plumb important, as well as pleasin' to the ear, and he didn't want to miss a word of it. So would she mind speaking a little louder?

Likewise, in this passage recounting a late-night telephone exchange from "No Room in My Heart to Forgive" (P18), we find a first person narrator revealing reasoning and thinking:

> "You scared me half to death", I said shakily. "What's wrong"?
>
> "Janice, nobody answers at the apartment"! Her voice came shrill. "I'm absolutely frantic! That stupid girl might have gone off and left Francie"!
>
> "Oh, she wouldn't do that", I said. "She's probably fallen asleep and doesn't hear the phone. But if you're worried you can go home and check "—

> "I can't leave the party! We're at Ken Thom's apartment, and when one couple leaves early everything falls flat! Old Mr. Thom is already down on Wally, and we simply can't afford to get Ken mad at us"—I was all set for what came next. "Janice, could you possibly go over and make sure everything's all right? I'll call you there in ten minutes"—
>
> "I can't make it in ten minutes"—Wondering, as I said it, why I should make it at all. Why should I go over at midnight to check on Francie, when her parents didn't care enough to leave a party?
>
> "Fifteen minutes, then! Please, Janice. I'll be glad to pay you"—So sure that money could do anything!

After careful consideration of the interactions between the rhetorical priming categories involved in this language dimension, I named it "Writing for the Intellect." This name distinguishes it from both the positive and negative features of the first language dimension and the name reflects the corpus texts' usage of this dimension to help readers follow reasoning and all them to get inside the text's argument.

5.3 Dimension 3: Retrospecting vs. Notifying

At first glance, the negative priming categories of the third factor in table 5.1—"notifying" and "direct activity"—would seem to overlap substantially with the negative categories of the first language dimension. How-

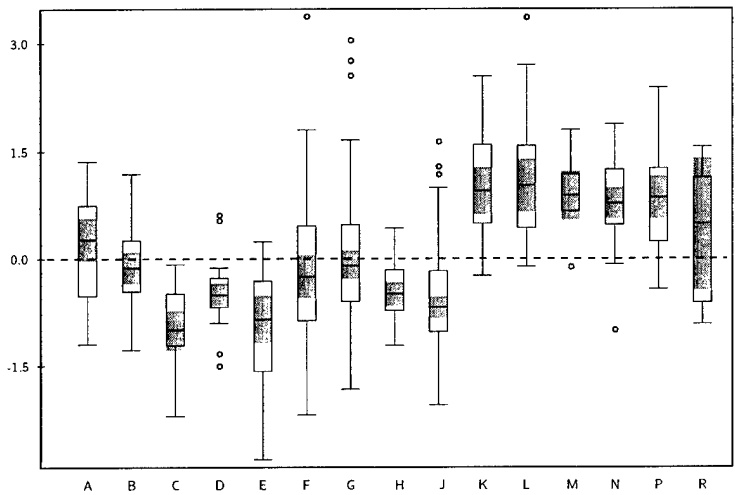
ever, on closer inspection it appears this language dimension captures perspective in the texts of the corpus.

Kaufer & Butler (2000, pp. 132–44) teach about using "eyewitnesses" to build accountings of a past, allowing a reader to experience an historical world through an eyewitness point of view. Kaufer and Butler set this notion of eyewitness accounting against texts that notify the reader about a world, reporting on events, but without the retrospection that only an eyewitness could provide.[5] The usage of this language dimension seems to capture this distinction in texts. The positive rhetorical priming categories are found in texts of the corpus in which the perspective is that of an eyewitness. When these priming categories are absent and "notifying" priming is present, the texts don't provide a retrospective accounting, but notify the reader about events and occurrences.

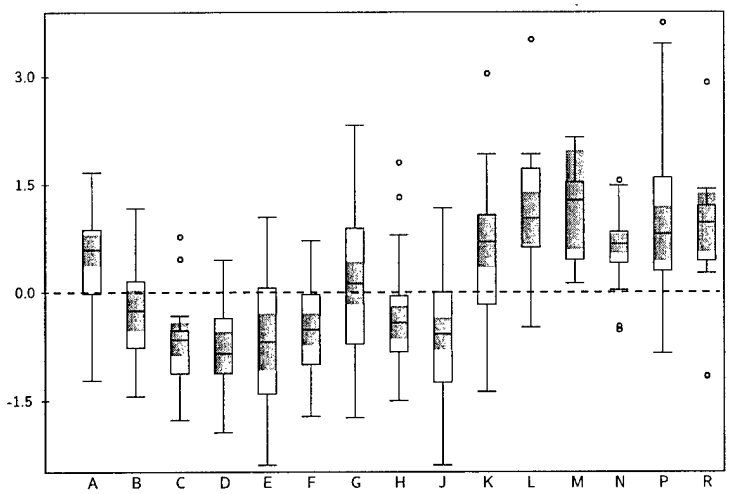
For example, in this text by Bonnie Prudden (E24) we find the writer relaying the past to us, but not as eyewitness. The writer here is not retrospecting on the past, she is relaying it, notifying us of its lessons and emphasizing aspects of the past she seems to want us to draw from her portrayal:

> The Russian gymnasts beat the tar out of the American gymnasts in the 1960 Olympics for one reason—they were better. They were better trained, better looking, better built, better disciplined... and something else—they were better dancers. Our athletes are only just beginning to learn that they must

[5]Of course this "eyewitness" could be fictional.



(a) Brown Corpus texts (pooled $SD = 0.79$)



(b) Frown Corpus texts (pooled $SD = 0.89$)

FIGURE 5.3: Boxplots showing dimension 3, "Retrospecting vs. Notifying," scores for the text genres

> study dance. The Russians are all trained as dancers before they start to study gymnastics. . . .
>
> The apparatus used by gymnasts was once a common sight in American gyms, but about 1930 it was dropped in favor of games. The parallel bars, horse, buck, springboard, horizontal bar, rings, and mats formerly in the school gyms were replaced by baseball, volleyball, basketball and football. But the Russians use gymnastics as the first step in training for all other sports because it provides training in every basic quality except one, endurance.

In this quotation the author uses "notifying" priming to signal that important concepts are being delivered by the text. Kaufer & Butler (2000) demonstrate that eyewitness are able to build credibility through retrospections of the events recounted. On the other hand, writers notifying their readers of events rely on other sources of credibility, as we saw in Prudden's quotation.

A further understanding is provided by examining the following passage by Julia Newman (E12) in which she opens an article by telling us of the sights to be experienced on a cross-country trip:

> THOSE WHO have never traveled the width and length of this land cannot conceive, on the basis of textbook description alone, the overwhelming space and variety of this country held together under one government. The miracle of democratic America comes home to one most strongly only when one has

> seen the endless Great Plains of the Midwest; the sky-reaching peaks of the Northwest mountains; the smoke-filled, art-filled, drama-filled life of the great cities of the East; the lush and historic charm of the South. Now, to add to the already unbelievable extremes found in one nation, we have the two new states of Hawaii and Alaska.

One might assume the writer has experienced all of this and one might expect the article will go on to provide more detail. But Newman breaks this early impression—over-riding any assumption of a retrospective accounting—by introducing "notifying" priming with a hint of market-speak in the last two sentences of her opening paragraph:

> To hope to cover just one region of this land and to enjoy all of its sights and events and, of course, to bring back pictures of your experiences, requires advance planning. For this reason, U.S. Camera has prepared this special U.S.A. vacation feature.

In doing so, Newman signals how the reader should take the remainder of her article (as a thinly-veiled marketing tool for a camera reseller).

In contrast, consider this passage by Mabel Wolfe Wheaton and LeGette Blythe (G54) that makes use of the positive priming categories of the language dimension—"think back," "past events," "linear guidance," and "shifting events":

> When Fred wheeled [my father] back into his room, the big one looking out on the back porch, and put him to bed, Papa told

> him he was very tired but that he had enjoyed greatly the trip downtown. "I've been cooped up so long", he added. Getting out again, seeing old friends, had given his spirits a lift.
>
> That night after supper I went back over to 48 Spruce Street—Ralph and I at that time were living at 168 Chestnut—and Ralph went with me. Papa was still elated over his afternoon visit downtown. "Baby, I saw a lot of old friends I hadn't seen in a long time", he told me, his eyes bright. "It was mighty good for the old man to get out again". The next day he seemed to be in fairly good shape and still in excellent spirits. But a few days after Fred's return he began hemorrhaging and that was the beginning of early and complete disintegration. It began in the morning, and very quickly the hemorrhage was a massive one. . . .
>
> They made the tests [for relative blood-match transfusion] and came to Fred; by now it was perhaps two days or longer after Papa had begun hemorrhaging. "Fred, your blood matches your father's, all right", Dr. Glenn said. "But we aren't going to let you give him any".
>
> "But why in the name of God can't I give my father blood"? Fred demanded. "Why can't I, Doctor"?
>
> "Because, Fred, it could do him no good. It's too late now. He's past helping. He's as good as gone".

Like other texts making use of this language dimension this text provides

a retrospective accounting of happenings. But such texts need not be written from the first person point of view.

For example, in this passage by Lillian Rogers Parks and Frances S. Leighton (G41) we find a retrospective perspective from a narrator who is not present in the scene she describes:

> All the rest of the days in the White House would be shadowed by the tragic loss, even though the President tried harder than ever to make his little dry jokes and to tease the people around him. A little boy came to give the President his personal condolences, and the President gave word that any little boy who wanted to see him was to be shown in. Backstairs, the maids cried a little over that, and the standing invitation was not mentioned to Mrs. Coolidge.
>
> The President was even more generous with the First Lady than he had been before the tragedy. He would bring her boxes of candy and other presents to coax a smile to her lips. He brought her shawls. Dresses were short in the days of Mrs. Coolidge, and Spanish shawls were thrown over them. He got her dozens of them. One shawl was so tremendous that she could not wear it, so she draped it over the banister on the second floor, and it hung over the stairway. The President used to look at it with a ghost of a smile.
>
> Mrs. Coolidge spent more time in her bedroom among her doll collection. She kept the dolls on the Lincoln bed. At night,

> when Mama would turn back the covers, she would have to take all the dolls off the bed and place them elsewhere for the night. Mama always felt that the collection symbolized Mrs. Coolidge's wish for a little girl. Among the dolls was one that meant very much to the First Lady, who would pick it up and look at it often.

Similarly, texts with a third person point of view that recount happenings score highly on this dimension. For example, this passage from Thomas B. Dewey's *Hunter at Large* (L06) gives the reader the feel of narrator as eyewitness, providing a retrospective account of happenings seen and heard:

> When his vision cleared he saw the taller one scrambling upward, reaching. Mickey was on his knees when Roberts turned on the stairs and the razor flashed in his hand. He felt his empty pocket and knew that Roberts had retrieved the only weapon at hand. Mickey's eyes fixed on the other's feet, which would first betray the moment and direction of an attack. He rose stiffly, forcing his knees to lock. The knifelike pain in his groin nearly brought him down again. He made himself back off slowly, his eyes wary on Roberts, who now had no more to lose than he. The pain dulled as he moved, and he steadied inside. After a moment he extended one hand, the fingers curled.
>
> "Come on", he said. "You want to be that big a fool—I was hoping for this". Roberts brushed at his eyes with his free hand

> and started down the steps. He held the razor well out to one side. He was invulnerable to attack, but he could be handled, Mickey knew, if he could be brought to make the first move.

Likewise, Richard Ferber's *Bitter Valley* (N05) gives the reader a detailed recounting, also providing an example of authors' usage of "linear guidance" priming in retrospection, used to help reader navigation through the text:

> She brought up her free hand to hit him, but this time he was quicker. He side-stepped her blow and she fell, stumbling against the gelding. She finally regained her balance and got up in the saddle. Her hat had come off and fallen behind her shoulders, held by the string, and he could see her face more clearly than he had at any time before. He had forgotten that she was so pretty. But her prettiness was what he had noticed first, and all the other things had come afterward: cruelty, meanness, self-will. He had known women like that, one woman in particular. And one had been too many. He watched the girl until she had gone into the trees, and waited until he couldn't hear the sound of her horse any longer, then went up to where the children were sleeping. They weren't sleeping, of course, but they thought they were doing him a favor by pretending.

By providing the "linear guidance" priming, the writer works to make the space of the text coherent and more easily navigable.

Likewise, the following passage by Arthur Miller (K25) also shows the co-occurrence of "shifting events" priming with the other dimensions as the writer moves the perspective easily between times of the narrative. The time intervals aren't carefully specified (which would be suggested by a use of "time interval" priming), but the shifts in time narrative time are signaled:

> He was in his mid-fifties at this time, long past the establishment of his name and the wish to be lionized yet once again, and it was almost a decade since he had sworn off lecturing. There was never a doubt any more how his structures would be received; it was always the same unqualified success now. He could no longer build anything, whether a private residence in his Pennsylvania county or a church in Brazil, without it being obvious that he had done it, and while here and there he was taken to task for again developing the same airy technique, they were such fanciful and sometimes even playful buildings that the public felt assured by its sense of recognition after a time, a quality of authentic uniqueness about them, which, once established by an artist as his private vision, is no longer disputable as to its other values.
>
> Stowey Rummel was internationally famous, a crafter of a genuine Americana in foreign eyes, an original designer whose inventive childishness with steel and concrete was made even more believably sincere by his personality. He had lived for al-

> most thirty years in this same stone farmhouse with the same wife, a remarkably childish thing in itself; he rose at half-past six every morning, made himself some French coffee, had his corn flakes and more coffee, smoked four cigarettes while reading last Sunday's Herald Tribune and yesterday's Pittsburgh Gazette, then put on his high-topped farmer's shoes and walked under a vine bower to his workshop.

In the examples presented so far I have focussed on the "imaginative" genres, which use the positive features of the language dimension more frequently (see figure 5.3(a) for confirmation). Some texts of the non-fiction genres also make use of the positive priming features to signal retrospection.

For example, in the "learned" text by Raymond J. Corsini et al., "Role-playing in Business and Industry" (J30), we find the co-occurrence of the positive priming features as the writers work to demonstrate the validity of their argument by recounting evidence from a series of interviews and case studies:

> In considering roleplaying for analysis we enter a more complex area, since we are now no longer dealing with a simple over-all decision but rather with the examination and evaluation of many elements seen in dynamic functioning. Some cases in evidence of the use of roleplaying for analysis may help explain the procedure. An engineer had been made the works manager of a firm, supplanting a retired employee who

> had been considered outstandingly successful. The engineer had more than seven years of experience in the firm, was well trained, was considered a hard worker, was respected by his fellow engineers for his technical competence and was regarded as a "comer".
>
> However, he turned out to be a complete failure in his new position. He seemed to antagonize everyone. Turnover rates of personnel went up, production dropped, and morale was visibly reduced. Despite the fact that he was regarded as an outstanding engineer, he seemed to be a very poor administrator, although no one quite knew what was wrong with him. . . .
>
> After the diagnosing, he left the course [administered by the authors], convinced that it could do him no good. We may say that his problem was diagnosed but that he refused treatment. The engineer turned works manager had a particular view of life—and refused to change it. We may say that his attitude was foolish, since he may have been a success had he learned some human relations skills; or we may say that his attitude was commendable, showing his independence of mind, in his refusal to adjust to the opinions of others. In any case, he refused to accept the implications of the analysis, that he needed to be made over.

Similar usage of the positive priming categories is sometimes found in accounts of scientific experiments in which the set up and findings are told

as a retrospective account. However, it's not typical of the "learned" genre (and the other "informative" genres) to find the usage of this language dimension's positive priming categories.

More typical of the "learned" texts, the following quotation by C.R. Wylie, Jr., (J21) suppresses indications of thinking and of "past events." Instead, as many language scholars interested in academic discourse might have predicted (e.g., Geisler, 1994), the typical texts of the "learned" genre use a present tense notification of scientific observations:

> In §1 we investigate a new series of line involutions in a projective space of three dimensions over the field of complex numbers. These are defined by a simple involutorial transformation of the points in which a general line meets a nonsingular quadric surface bearing a curve of symbol fmula. Then in §2 we show that any line involution with the properties that (a) It has no complex of invariant lines, and (b) Its singular lines form a complex consisting exclusively of the lines which meet a twisted curve, is necessarily of the type discussed in §1. No generalization of these results to spaces of more than three dimensions has so far been found possible.

We find a similar lack of the positive-priming features in the following "learned" passage by Kenneth Hoffman and Ray Kunze (J18). In this passage we also find usage of "notifying" priming as the authors attempt to help the reader process the dense strings of text that make up their conclusion:

> From these results, one sees that the study of linear operators on vector spaces over an algebraically closed field is essentially reduced to the study of nilpotent operators. For vector spaces over non-algebraically closed fields, we still need to find some substitute for characteristic values and vectors. It is a very interesting fact that these two problems can be handled simultaneously and this is what we shall do in the next chapter. In concluding this section, we should like to give an example which illustrates some of the ideas of the primary decomposition theorem. We have chosen to give it at the end of the section since it deals with differential equations and thus is not purely linear algebra.

In addition to "notifying," this quotation also has examples of "direct activity" priming, a co-occurring negative priming feature. The association between "notifying" and "direct activity" is easily seen in instructive texts, such as this passage by Ann Carnahan (E14) in which she guides readers through the "do's and don'ts" of cooking out:

> Keep ashes from one barbecue to the next to sprinkle over coals if they are too hot, and to stop flames that arise from melting grease.
>
> Do line barbecue fire bowl with heavy foil to reflect heat.
>
> Don't forget to buy a plastic pastry brush for basting with sauces. Clean it meticulously in boiling water and detergent, rinse thoroughly.

> Do build a wall of glowing coals six to eight inches in front of meat that is barbecued on an electric spit. Make use of the back of the barbecue or of the hood for heating vegetables, sauces and such.
>
> Don't fail to shorten cooking time by the use of aluminum foil cut slightly larger than the surface of steaks and chops. Sear on both sides then cover meat loosely with heat reflecting foil for juiciest results.

In the texts of the "skills and hobbies" genre, the co-occurrence of "notifying" and "direct activity" priming features appears strong. However, this co-occurrence otherwise appears to be artifactual. In most of the corpus "direct activity" usage is found in texts that are retrospective, for example in texts where characters direct each other in carrying out actions. Furthermore, some "skills and hobby" texts combine a retrospective description of what the narrator did with instructions for the reader about what to do. As you will see, both of these cases are better explained by the fourth language dimension, described in the next section, where "direct activity" priming is salient.

5.4 Dimension 4: Instructing

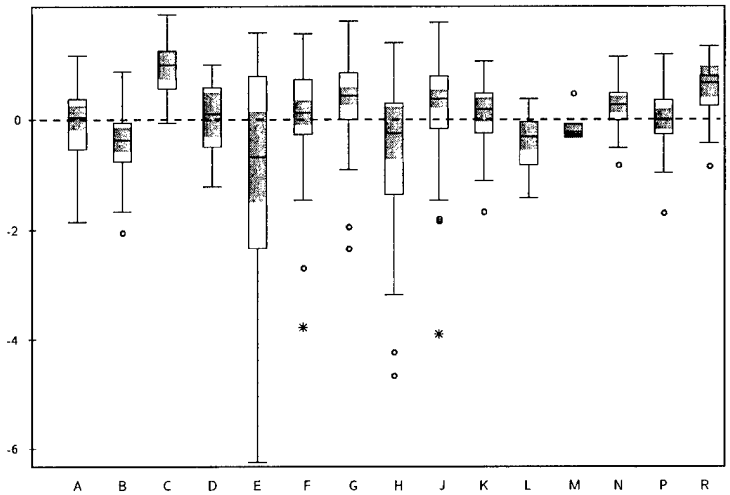
As you can see from table 5.1, all of the salient priming categories of the fourth and fifth language dimensions are negative. This means those categories were found to co-occur in the texts of the corpus; this does not

indicate that less of those negative priming categories are present in the language dimension (as is a common misconception).

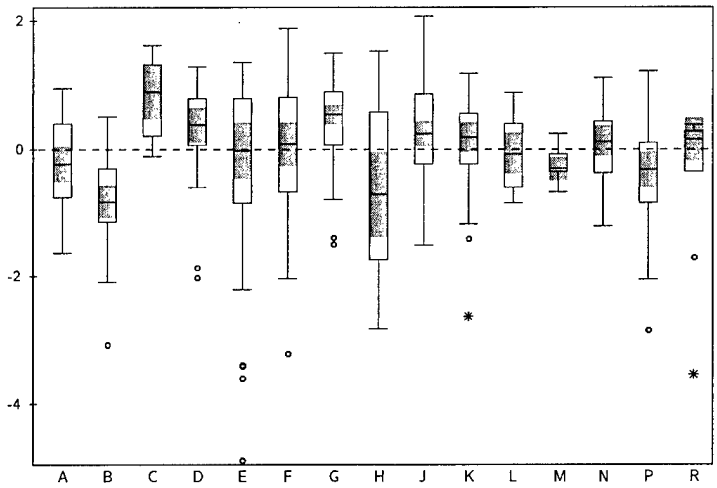
The salient text features of the fourth language dimension, "think ahead" and "direct activity," are found in texts of the corpus that provide instructions, projecting into a future that will be created (or at least affected) when the instructions are followed. For example, in this passage from a book titled *Organic Gardening and Farming* (E02), the author anticipates what the reader may expect, suggesting the actions to follow to enact that projected future:

> The day will come, in midsummer, when you find your plants becoming "leggy", running to tall-growing foliage at the expense of blossoms. Try pegging down each separate branch to the earth, using a bobby pin to hold it there. Pick the flowers, keep the soil dampened, and each of the pegged-down branches will take root and become a little plant and go on blooming for the rest of the season....
>
> Pansies have character. They stick to their principles, insist upon their due, but grow and bloom with dependable regularity if given it. Treat them right and they'll make a showing every month in the year except the frigid ones. Give them food, some shade, mulch, water and more food, and they'll repay your solicitude with beauty.

Later in the same text we find another passage with similar co-occurrence of the "think ahead" and "direct activity" priming features:



(a) Brown Corpus texts (pooled $SD = 0.92$)



(b) Frown Corpus texts (pooled $SD = 0.89$)

FIGURE 5.4: Boxplots showing dimension 4, "Instructing," scores for the text genres

> A SALAD WITH greens and tomato is a popular and wonderfully healthful addition to a meal, but add an avocado and you have something really special.
>
> This delightful tropical fruit has become well-known in the past thirty years because modern transportation methods have made it possible to ship avocado anywhere in the United States. It has a great many assets to recommend it and if you haven't made avocado a part of your diet yet, you really should. You will find that avocado is unlike any other fruit you have ever tasted.
>
> It is roughly shaped like a large pear, and when properly ripened, its dark green skin covers a meaty, melon-like pulp that has about the consistency of a ripe Bartlett pear, but oily. The avocado should have a "give" to it, as you hold it, when it is ripe. The flavor is neither sweet, like a pear, nor tart like an orange; it is subtle and rather bland, nut-like. It is a flavor that might take a little getting used to—not because it is unpleasant, but because the flavor is hard to define in the light of our experience with other fruits. Sometimes it takes several "eatings" of avocado to catch that delightful quality in taste that has made it such a favorite throughout the world. Once you become an avocado fan, you will look forward to the season each year with eager anticipation

The writer of these quotations—one passage helping us achieve healthy

pansies and the second helping us pick ripe avocados—both direct us in our motions and in a proximate environment. This is not necessarily true, however, of all texts making use of this language dimension.

For example, the anonymous author of this text uses the priming dimensions of the language dimension. similarly, projecting ahead into the future and suggesting activity to effect (or avoid) specific outcome, but the task is not a physical one:

> Have the insurance company or your own accounting department break down the cost of your insurance package periodically. You may find certain coverage costing much more than is economically feasible, thereby alerting you to desirable revisions. Check to see if some of your benefits—such as on-the-job disability pay—can be put on a direct payment rather than an insured basis at a savings to you. Use deductable insurance wherever feasible. It can put an end to marginal claims which play havoc with your insurance rates.
>
> Also, beware of open-end policies, especially in the medical field. This will mean that every time there's an increase in hospital rates your cost will go up in like manner. Put a dollar-and-cents limit on benefits. Don't go overboard on insurance that pays benefits only upon death. Generally, your employee will greatly appreciate benefits that protect him during his working life or during retirement.

Notice, also, the presence of the third co-occurring priming category from

table 5.1, "interacting."

In each of the above examples, the writer's directives are aimed at the reader. However, the "interacting" priming category was found not to be salient to this language dimension because the usage of the language dimension is often found in texts that are not directing readers to action but are recommending action in general or are capturing interactions between characters in fictional works.

For example, the following passage by Bonnie Prudden (E24) urges American society to overcome a felt difficulty through the installation of chin-up bars at the classroom door:

> President Kennedy has asked that we become a physically fit nation. If we wait until children are in junior high or high school, we will never manage it. To be fit, one has to start early with young children, and today the only person who really reaches such children is the teacher of dance. If the dance teachers of America make it their business to prepare their young charges for the gymnastics that must come some day if our schools are really responsible, we will be that much ahead. School teachers, all too unprepared for the job they must do, will need demonstrators. . . .
>
> For arm and shoulder strength a chinning bar is recommended. It should be installed over a door that is in full view of everyone, and a chair should be placed under it, a little to one side. Those children who can chin themselves should be told

> to do one chin up each time they pass under it. Those who are too weak, should climb on the chair and, starting at the top of the chin, let themselves slowly down. When they can take ten seconds to accomplish the descent, they will have the strength to chin up. Parents should be informed about this system and encouraged to do the same with the whole family at home.

As urgent as Prudden's plea seems, it's not entirely clear to whom this text is addressed—perhaps her directions are meant for parents or for school administrators? The "interacting" priming here is different than that of the previous passages.

This difference is seen even more clearly in the following example from the conclusion of a political science "learned" essay (J22). Here again notice the author uses "think ahead" and "direct activity" priming not to suggest the *reader* should jump up and do something, but suggesting a potential future the government might affect:

> The United States can help by communicating a genuine concern with the problems these [developing] countries face and a readiness to provide technical and other appropriate forms of assistance where possible. Our central goal should be to provide the greatest positive incentive for these societies to tackle boldly the tasks which they face.
>
> At the same time, we should recognize that the obstacles to change and the lack of cohesion and stability which characterize these countries may make them particularly prone to diver-

sions and external adventures of all sorts. It may seem to some of them that success can be purchased much less dearly by fishing in the murky waters of international politics than by facing up to the intractable tasks at home. We should do what we can to discourage this conclusion, both by offering assistance for their domestic needs and by reacting firmly to irresponsible actions on the world scene. When necessary, we should make it clear that countries which choose to derive marginal advantages from the cold war or to exploit their potential for disrupting the security of the world will not only lose our sympathy but also risk their own prospects for orderly development. As a nation, we feel an obligation to assist other countries in their development; but this obligation pertains only to countries which are honestly seeking to become responsible members of a stable and forward-moving world community. . . .

American policy should press constantly the view that until these governments demand efficiency and effectiveness of their bureaucracies there is not the slightest hope that they will either modernize of democratize their societies. We should spread the view that planning and national development are serious matters which call for effort as well as enthusiasm. Above all, we should seek to encourage the leaders of these societies to accept the unpleasant fact that they are responsible for their fates.

And again in the following text by Thomas D. McGrath (J71) we find a typical usage of this language dimension with the author providing a short example of what *might* result (or be avoided) through the recommended action although, again, the action doesn't appear to be one a reader is intended to enact:

> Antisubmarine warfare does not involve clashes between large opposing forces, with the decision a result of a single battle. It is a war of attrition, of single actions, of an exchange of losses. This exchange must result in our ending up with some effective units. Initially, having fewer units of some elements—especially submarines—than the opponent, our capabilities need to be sufficiently greater than theirs, so that the exchange will be in our favor. Therefore, the third principle of the plan must be that it does not depend for effectiveness on engagement by the same types, unless at an assured favorable exchange rate.

The genre making the least use of this language dimension are the texts of the "newspaper reviews" genre. These texts provide little "direct activity" and "think ahead" priming. In the few "newspaper review" texts that do include the salient priming categories, the effects are usually residual, coming, for example, from a quoted passage from a book being reviewed) or are found at the periphery of the review itself as in this example from *The Providence Journal* (C10) in which the writer projects ahead in directing an interested reader to buy the album:

> Kid Ory, the trombonist chicken farmer, is also one of the solid

> anchor points of jazz. He dates back to the days before the first sailing ship pulled into New Orleans. His horn has blown loud and clear across the land for more years than he cares to remember. Good Time Jazz has released a nice two-record album which he made.... All this will serve to show off the Ory style in fine fashion and is a must for those who want to collect elements of the old-time jazz before it is too late to lay hands on the gems.

Likewise, although most texts of the "belles lettres, biography, memoir" genre in the corpus do not include much usage of this language dimension. The few texts that do make use of the language dimension do so when dispensing advice or drawing lessons for the reader. For example, the essay by Kenneth Reiner, "Coping with Runaway Technology" (G22), is unusually directive for its genre:

> Properly mindful of all the cultures in existence today throughout the world, we must employ these resources without war or violent revolution. If we were creating a wholly new society, we could insist that our social, political, economic and philosophic institutions foster rather than hamper man; best growth. But we cannot start off with a clean slate. So we must first analyze our present institutions with respect to the effect of each on man's major needs. Asked which institution most needs correction, I would say the corporation as it exists in America today.

I've named this language dimension "overt instructions." Its usage is most typical of the "skills and hobbies" texts of the corpora. Kaufer & Butler (2000, chapter 6) write about "invitations to do and learn." Their description of the rhetorical priming features involved in these "invitations" has substantial overlaps with this language dimension; however, the distinction between Kaufer and Butler's term, "invitation," and my term, "instruction," is significant, although not immutable, as those who've seen a James Bond villain "invite" Mr. Bond to dine can attest.

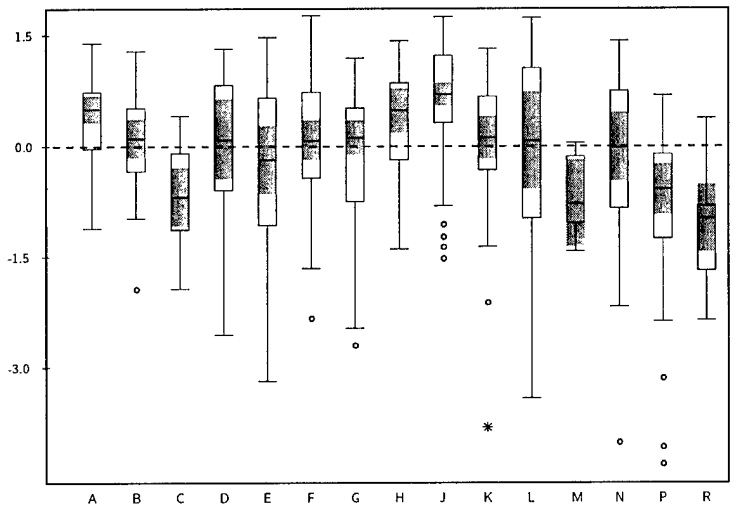
Judging from the co-occurrences of the priming features in the Brown corpus, "instruction" is more apt as a description of what the 1961 writers did. However, comparing the usage of this language dimension between the Brown and the Frown corpora, it seems the "skills and hobbies" genre changed markedly (see figure 5.4, p. 179). Perhaps this reflects a change toward a more "invitational" and less "instructive" writing stance in contemporary technical writing, which was the basis of Kaufer and Butler's original insight.

5.5 Dimension 5: Referencing Positive Relationships

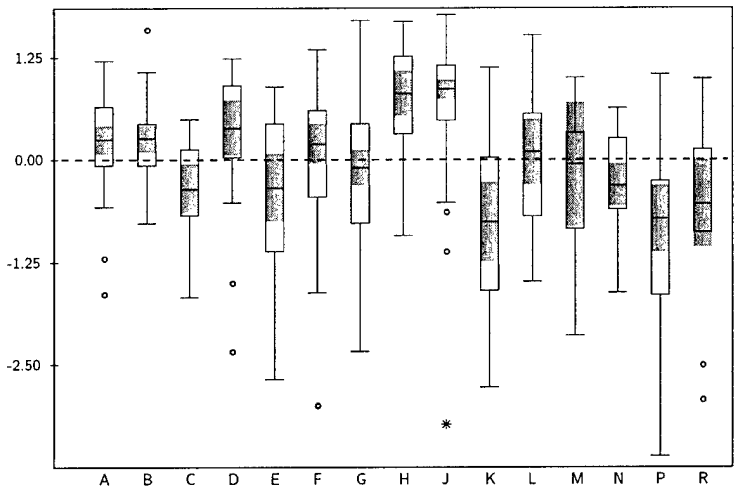
One of the ways it seems the instruction writers of the Brown corpus softened their instructions is by using the final language dimension, which I've named "Referencing Positive Relationships." For example, the writer of the text from *Organic Gardening and Farming* (E02) continues the discussion of pansies with the following:

> Nothing is easier to grow from seed than pansies. They germinate quickly, the tiny plants appearing in a week, and grow along lustily. It doesn't really matter which month of the year you sow them, but they germinate best when they have a wide variation of temperature, very warm followed by cool in the same 24 hours. I like to make a seedbed right in the open, though many people start them successfully in cold frames.
>
> Pansies don't have to be coddled; they'd rather have things rugged, with only moderate protection on the coldest days. If you do use a cold frame be sure that its ventilation is adequate. For my seedbed I use good garden soil with a little sand added to encourage rooting. I dig it, rake it smooth, sow the seeds and wet them down with a fog spray. Then I cover the sowing with a board. This keeps it cool and moist and protects it from birds. Ants carry away the seeds so better be sure that there are no ant hills nearby.

This text exemplifies two of the co-occurring categories of priming—"first person" and "think positive"—that dominate this language dimension, both with factor loading scores with magnitudes over 0.65 (see table 5.1). In addition, a third co-occurring though weaker priming category, "interacting," is at work in the above quotation: the text seems to give a reader assurances and some motivation. As we read, the author seems to interact in sharing positive thoughts about the experience. This usage of the language dimension is typical for texts of the "skills and hobbies" genre.



(a) Brown Corpus texts (pooled $SD = 0.92$)



(b) Frown Corpus texts (pooled $SD = 0.82$)

FIGURE 5.5: Boxplots showing dimension 5, "Referencing Positive Relationships," scores for the text genres

Another place this language dimension's rhetorical priming features co-occur is in works that attempt to build an almost conspiratorial relationship between the narrator and reader. Take, for example, the following passage from Peter Bains' essay "With Women... Education Pays Off" (N18) from the "adventure and western" genre. In the passage notice the narrator's asides to the reader and the usage of direct address, "interacting" rhetorical priming that helps the reader sense a relationship:

> If it were not for an old professor who made me read the classics I would have been stymied on what to do, and now I understand why they are classics; those who wrote them knew people and what made people tick. I worked for my Uncle (an Uncle by marriage so you will not think this has a mild undercurrent of incest) who ran one of those antique shops in New Orleans' Vieux Carre, the old French Quarter. The arrangement I had with him was to work four hours a day. The rest of the time I devoted to painting or to those other activities a young and healthy man just out of college finds interesting....
>
> I loved the city and I particularly loved the gaiety and spirit of Mardi Gras. I had seen two of them and we would soon be in another city-wide, joyous celebration with romance in the air; and, when you took a walk you never knew what adventure or pair of sparkling eyes were waiting around the next corner. The very faces of the people bore this expectation of fun and pleasure. It was as if they could hardly wait to get into their

> costumes, cover their faces with masks and go adventuring.

Other "imaginative" texts of the corpus use a more subtle form of "interacting" priming, asking questions to reveal narrator or character thinking. We saw this usage earlier in Wesley Newton's passage about the Japanese soldier in World War II (in section 5.1). Additionally, some "imaginative" texts make use of this language dimension when revealing relationships among the characters of a work, using interactive dialog to show an existing (or budding) relationship. We see both usages of "interacting" priming (in addition to "first person" and "think positive" priming) in this passage by Marvin Schiller (K29):

> Was it love? I had no doubt that it was. During the rest of the summer my scholarly mania for making plaster casts and spatter prints of Catskill flowers and leaves was all but surpassed by the constantly renewed impressions of Jessica that my mind served up to me for contemplation and delight....
>
> I asked Jessie to spend New Year's Eve with me. Lovingly, she accepted, and so great was my emotion that all I could think of saying was, "You're amazing, you know"? Later, we agreed to think of how we wished to spend that night. We would write to one another and make a definite plan....
>
> It would be the first time I had ever been completely alone with a girl I loved. I had no idea of what subjects one discussed when alone with a girl, or how one behaved: Should I hold her hand while walking or only when crossing the street? Should

> I bring along a corsage or send one to her? Was it preferable to meet her at home or in the city? Should I accompany her to the door of her home, or should I ask to be invited in? In or out, should I kiss her goodnight?

This text is unusual for the "general fiction" genre. Most of the texts of that genre do not make such extensive use of the priming features of this language dimension. Schiller's usage is more typical of the texts in the "romance and love story" genre in which the three priming categories of this language dimension quite commonly co-occur (as you can see in figure 5.5(a)).

For example, this passage from Spencer Norris's essay, "Dirty Dog Inn" (P20), reveals the narrator's private, generally positive thoughts to the reader and interacts with other characters through dialog:

> I was slowly swimming down to the bottom of the sea. She made me welcome. Her dark cool caresses were sweeter than any woman's; the many little tricks she knew made her embrace the ultimate one—the ever more fantastic pressures deeper in her body squeezed not me but the air I breathed into a nitrogen anesthetic.
>
> Yielding-Mediterranean-woman-flesh-of-water, she soothed me, and drew me deeper into her. I no longer knew how deep I was, somewhere under 230 feet, getting drunker, happier and more contented by the second. The reasons for this dive seemed foolish now. Only the dive itself had any meaning. . . .

While a hazy part of my mind concentrated on swimming down, a clear part sorted over recent events, among them my only positive act in a long time. It was when I packed up what duds I had and went to Paris. It was no vacation, just me getting out after a bellyfull. I knew it wouldn't be the same. Wild kicks never are, but I hoped to dig up a better frame of mind. Once before I had been to Paris, long before I married Valery. That first time was good and it stuck with me....

I was no yokel, but I was young, and this was Paris! I had champagne at Maxim's, then went into a cafe called the Jour et Nuit to ask the way to Montmartre. I never got there. I met Claire, which was better.

She was eating bread and cheese just as fast as she possibly could, and washing it down with red wine. I stared. I didn't know a human could feed so fast and still be beautiful. She was blonde, and young, and nice and round in a tight white dress. Maybe her ravenous eating wasn't grotesque because she was so positive about it. When she had drained the last of the bottle and paid her bill, she came directly to my table and said: "Handsome soldier, I have assuaged one hunger with food. I feel another of terrible urgency. Is your evening free"?

"Madame", I said with noblesse oblige because of the "handsome" —"yeah". And so off we went to her apartment.

The first part of this passage, as the reader is moved from swimming in the Mediterranean to view a previous experience in Paris, makes use of the fourth co-occurring priming category of this language dimension, "shifting events." This priming category is relatively less powerful (with a factor loading score of -0.40) than the other three categories and is not consistently used by authors of the corpus showing relationships. However, when "shifting events" is present, it seems to heighten the feeling of shared experience available from the text.

For example, in the text "A Husband Stealer from Way Back" (P22), we grow up with the narrator as she reflects back on past experiences to relay what led her to love the cheating boy friend (who becomes husband). As in most fiction texts with more than one character, we find the "interacting" priming between the characters of the story and in addition to between the narrator and reader. This priming seems intended to nudge the reader to empathize with the narrator's long-term dilemma:

> I was really upset the first time I discovered that my boy friend Johnnie was seeing Mrs. Warren. I asked him about it one night while we were sitting in his truck. I asked him if it was true. He gave me a straight, honest answer. "Look, Sue baby", he'd said. Much as I love you—well, a guy's a guy and Lucille's willing to—to come across. Honest, kitten, that's all it is—I don't even like Lucille much".
>
> I guess it was at that moment that I realized what I was up against in the person of Lucille Warren. But it didn't seem fair.

> My love for Johnnie was young and clean—how could I possibly compete with a woman like that, who didn't hesitate to use her sex. Johnnie was a trucker with a small lumber outfit in a town about twenty miles away, and he was also pretty good at anything in the carpentry line.

Similar patterns of usage of this language dimension are found in some of the "informative" genres. For example, this passage from a travel article by Julia Newman (E12) interacts with the reader, suggesting positive benefits as the reader is shifted along a virtual tour of the northeast:

> . . . the Northeast offers historic battlefields; lovely old villages and a rugged seashore among its many worthwhile sights. The rolling farms of Maryland, the peerless metropolis of New York City, the verdant mountains of Vermont can all be included in your Northeast vacation. By automobile from New York, for example, you can take a one or two-day tour to Annapolis, Maryland to see the colonial homes and the U.S. Naval Academy (where you can shoot [pictures of] the dress parade on Wednesdays); to Washington, D.C., for an eye-filling tour of the city; or to Lancaster, Pa. , the center of the Pennsylvania Dutch country; Philadelphia with its historic buildings and nearby Valley Forge; to West Point, N.Y., the famous military academy in a beautiful setting on the Hudson River.
>
> New England deserves as much of your vacation time as you can afford with such areas as Cape Cod providing won-

> derful beaches, artists' colonies and quaint townships. From here you can easily include a side trip to the old whaling port of Nantucket, Massachusetts, which looks just as it must have two centuries ago.

The texts of the "newspaper reviews" genre also make frequent use of this language dimension, but its use tends to emphasize "think positive" more than the other co-occurring priming categories. For example, this excerpt of a review from *The New York Times* (C03) provides "interacting" priming to accompany the positive review:

> THE Polish song and dance company called Mazowsze, after the region of Poland, where it has its headquarters, opened a three-week engagement at the City Center last night. A thoroughly ingratiating company it is, and when the final curtain falls you may suddenly realize that you have been sitting with a broad grin on your face all evening. Not that it is all funny, by any means, though some of it is definitely so, but simply that the dancers are young and handsome, high-spirited and communicative, and the program itself is as vivacious as it is varied. There is no use at all in trying to follow it dance by dance and title by title, for it has a kind of nonstop format, and moves along in an admirable continuity that demands no pauses for identification.

In the "learned" texts of the corpus, in general, we don't consistently find co-occurrences of the salient priming categories of the language di-

mension. Harry H. Hull's article, "The Normal Forces and Their Thermodynamic Significance" (J03), provides a typical passage for the "learned" genre, making no moves to evoke a relationship with the reader in this (or any of the other) passage:

> A band viscometer is shown in Figure 2. It consists of two blocks with flat surfaces held apart by shims. There is a small well in the top in which the fluid or paste to be tested is placed. A tape of cellulose acetate is pulled between the blocks and the tape pulls the fluid or paste with it between the parallel faces of the blocks. In normal use weights are hung on the end of the tape and allowed to pull the tape and the material to be tested between the blocks. After it has reached terminal velocity, the time for the tape to travel a known distance is recorded. By the use of various weights, data for a force-rate of shear graph can be obtained. The instrument used for this work was a slight modification of that previously described.

Perhaps academic authors of such passages safely assume their readers will already be part of a community with built-in interest in their subject matter.

Nonetheless, there are exceptions in the texts of the "learned" genre, such as these passages excerpted from an open letter by Jimmy Ernst to artists in the Soviet Union (J62):

> Those whom I wish to address with this letter are for the most part unknown to me. It may well be that, when Rudy Poz-

> zatti and I visited your country last spring, you were living and working close to the places we saw and the streets we walked. As American artists, it was natural that we would want to meet as many Soviet artists as possible....
>
> I am very pleased that quite a number of you found ways to communicate to me your desire to hear of our reactions and experiences in the U.S.S.R. I can well understand your curiosity. We, ourselves, are always eager to know how others feel about us and the way in which we live. it is my hope that this written message and report will reach you through the good offices of the Union of Soviet Artists.

Ernst and the other authors of the corpus who make use of this final language dimension seem to be attempting to forge or, at least, reference positive relationships in their texts.

5.6 A meso-level rhetorical theory

The five language dimensions described in this chapter comprise a meso-level rhetorical theory, suggestive of the ways the authors of the texts in the corpora met their rhetorical challenges. The authors met these varied challenges, in part, by altering the language choices described by Kaufer et al.'s (In Press) micro-rhetorical priming categories. The authors altered their usages of these categories in consistent ways that were interpretted and described as five language dimensions:

Language Dimension 1 *Writing for the Eyes vs. Informing*

Language Dimension 2 *Writing for the Intellect*

Language Dimension 3 *Retrospecting vs. Notifying*

Language Dimension 4 *Instructing*

Language Dimension 5 *Referencing Positive Relationships*

The evidence suggests writers of the Brown and Frown corpora controlled for these language dimensions in effecting their textual designs. The next chapter takes up the understandings such a meso-level rhetorical theory of writing might afford.

6 Implications and Future Directions

The study presented in this book sought a significant relationship between micro-rhetorical priming features (Kaufer et al., In Press) and high-level rhetorical plans. This relationship was found using two 500-text corpora of published American English (Francis & Kucera, 1979; Hundt et al., 1999). These texts, composed some thirty years apart, had been sorted into 15 text genres by the corpus compilers. The corpora were analyzed to see whether differences in text genre accounted for co-occurrences of the micro-rhetorical priming features.

Five groups of co-occurring priming features were found to be significantly affected by text genre (at the $\alpha = 0.001$ or better level of significance) and were confirmed across both corpora, demonstrating authors' consistent usage of the text features according to genre. These groups of priming features were interpreted as language dimensions based upon a qualitative analysis of their usage in the corpora. These five language dimensions partially describe the differences in the ways the authors met

their rhetorical challenges across the text genres. The implication is that these five dimensions may mark fundamental rhetorical "cut points" in written English, functioning as a heretofore hidden meso-layer linking some micro-level linguistic decisions and authors' macro-level rhetorical plans and tasks. For example, when an author decides the rhetorical situation calls for her text to be highly descriptive, she will likely choose priming features of the "writing for the eyes" language dimension. Likewise, an author intending to teach an audience how to change the alternator on a car will likely choose priming features of the "instructing" language dimension.

The primary benefit of this finding for language scholars is that it makes available a set of pre-defined linkages between the language dimensions' functions (described in chapter 5) and the micro-level rhetorical priming features authors and speakers use to accomplish these functions. This chapter discusses several of the potential implications and research possibilities afforded by these findings.

6.1 An instrumental organization of rhetorical priming categories

Kaufer et al.'s (In Press) hierarchy of rhetorical priming categories (shown in figure 4.1, p. 91) represents a conceptual hierarchy of the categories. In this scheme the categories are logically ordered and the offered hierarchy helps to explain the authors' conceptualization of their language theory.

The findings of the current study suggest a different, instrumental

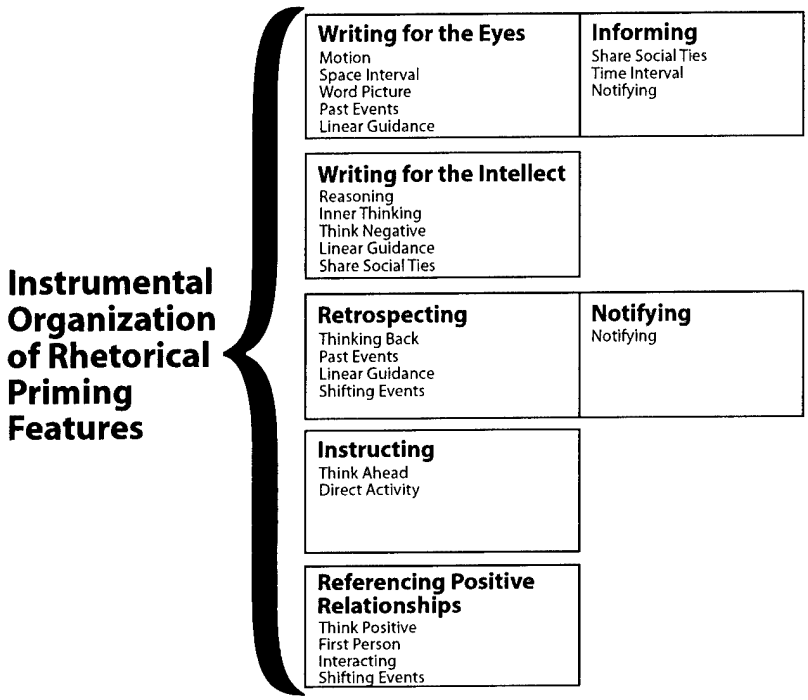


FIGURE 6.1: The micro-rhetorical priming features (Kaufer et al., In Press), grouped by their co-occurrences of usage within the Brown and Frown corpora texts.

ordering of the priming categories, depicted in figure 6.1. This might be considered a task-oriented hierarchy: the categories in this scheme are ordered according to their co-occurrences of usage in the studied corpora. This scheme reflects how the corpora writers used the rhetorical priming features in attempting to accomplish their writing goals.

Such a task-oriented conceptualization of the language theory provides for an understanding of the five language dimensions as latent indicators of authors' broad rhetorical strategies. They are "latent" in the

sense that each language dimension represents a broad decision about how the text is to affect its reader, but these decisions would probably elude direct or exact measurement (see Bollen, 2001, 2002). Nonetheless, the current study has demonstrated indicators of these decisions may be found by measuring the micro-rhetorical priming categories and using the co-occurrences shown in figure 6.1 as indications of what is happening rhetorically in an author's text.

As an example of how this latent analysis operates in practice, let's consider William Shakespeare's romance and history plays. One would expect Shakespeare to develop stronger, more overtly positive relationships between characters in his romance plays than he would in his histories. Indeed, when one tags and analyzes these two groups of plays (see Hylton, 2000, for electronic versions of the texts), one finds significant differences[1] in the two genres' usage of the priming features making up language dimension 5, "Referencing Positive Relationships." Shakespeare's usage of these priming features are concentrated in the romances. He uses this particular combination of features almost twice as often in the romances as he does in the history plays, on average. This suggests that Shakespeare—like the authors of the Brown and Frown corpora writing some 400 years after him—builds an audience's awareness of romance into texts by using the particular combination of priming features described by language dimension 5.

As another example of how the language dimensions serve as latent

[1]ANOVA results in $F(1, 21) = 12.42$, $p = 0.002$, for dimension 5.

indicators of underlying rhetorical strategies, let's reconsider the 1940s and 1950s tobacco advertisements that introduced the study (see chapter 1). The two decades of collected ads differ significantly[2] on their usages of the priming categories making up language dimensions 3 and 5. On language dimension 3, "Retrospecting vs. Notifying," the 1940s texts tend to provide a more retrospective stance, describing happenings as a direct observer might. On the other hand, the 1950s texts tend, instead, to notify the audience of occurrences that had taken place, without retrospection. Furthermore, the 1940s texts tended to use less of the priming categories making up the 5th language dimension, "Referencing Positive Relationships," as would be expected based upon the difference in wartime and peacetime subject matter.

Further qualitative analysis indicates the differences in groups of texts may be pointing toward the writers' need to respond to an altered notion of what constituted "progress" during these two decades of American history. To see this in action, let's contrast the following two quotations. The first is from a 1950s text. Here the text seems to attempt to engender a clear sense of positive scientific progress for the reader:

> Only Chesterfield is Made the Modern way—with AccuRay. Discover for yourself what modern science can do to increase your cigarette enjoyment... You'll marvel at the extra flavor that comes through. Yet because the measurably better

[2]Comparison using ANOVA shows differences on language dimension 3, $F(1, 49) = 3.72$, $p = 0.059$, and language dimension 5, $F(1, 49) = 4.01$, $p = 0.051$.

> cigarette smokes more slowly you enjoy a cool mildness never possible before.

Contrast that quotation with the following from a 1940s advertisement, which also seems aimed toward a scientific register, but excludes dimension 5 usage in that it dampens "think positive" priming and, instead, primes "think negative" in regard to aspects of tobacco usage and science's role in circumventing these negative effects. This 1940s quotation tells its story of scientific discovery in the past tense, relaying a retrospective viewpoint that has witnessed problems with tobacco usage and has seen science overcome them:

> When you inhale—and all smokers do, some of the time—there's increased chance of irritation. So, be guided by the findings of eminent doctors. They discovered: On comparing the irritant quality in the smoke of the four other leading brands was found to average more than three times that of the strikingly contrasted Philip Morris and the irritation lasts more than five times as long! Philip Morris brings you the delightful flavor and aroma of the world's finest tobaccos—with never a worry about throat irritation.

As you can see, by comparing the two groups' usages of the rhetorical priming features and looking at the features' co-occurrences, we find latent indications of some of the decisions authors may have made to accommodate the rhetorical situations of their texts.

Of course to fully substantiate this claim, further study of real-time discourse production would be necessary. My aim in introducing these examples from across the period of modern English is to provide further evidence of the stability of the theory and to indicate the powerful insights it might afford on language usage and genre analysis, a topic that is taken up further in the next two sections.

6.2 Latent indicators for genre analysis and text categorization

Among the most intriguing findings of this study are some of the changes seen in individual text genres between the two corpora. Although a full exploration of these findings is beyond the scope of this book, a discussion of some of the possibilities is worth consideration here. These findings tend to support contemporary notions of "genre" as an agent of both stability and change in writing (see, e.g., Coe et al., 2002) and are suggestive of a new, potentially fruitful method of genre analysis.

As an example, let's reconsider the texts of the "science fiction" genre in the corpora. Given the three decades that separate the composition of the two corpora, we might reasonably expect to find changes in this genre based, at least in part, upon scientific occurrences. The Brown corpus's 1961 texts were composed well before the lunar landings while the Frown corpus texts were all written after the 1986 space shuttle Challenger disaster and after inklings of a popular cyberspace and widespread nanotechnology were commonplace.

TABLE 6.1: Summary of the language dimension scores of the "science fiction" texts in each corpus

	Brown Corpus			Frown Corpus		
Dimension	mean	med.	*SD*	mean	med.	*SD*
1	0.153	0.131	0.656	0.520	0.498	0.877
2	0.839	0.393	0.867	0.694	0.916	0.584
3	0.889	0.886	0.633	1.146	1.288	0.741
4	-0.096	-0.213	0.291	-0.234	-0.294	0.305
5	-0.680	-0.771	0.555	-0.284	-0.050	1.086

$n = 6$; ANOVA showed no significant differences between groups at the $\alpha = 0.10$ level on any language dimension

As we saw in the examples presented in the last section, any consistent changes in the authors' usages of the rhetorical priming categories would affect the genre's typical scores on the language dimensions. Such changes, I have argued, are latent indications of the authors' underlying rhetorical strategies.

We find such changes in the "science fiction" genre's scores between the corpora, as is shown in table 6.1. However, none of the changes in the mean scores between the corpora is statistically significant at even the $\alpha = 0.20$ level. This means that once we take account of the variance among the groups of texts (using the ANOVA procedure) the differences between the groups cannot be attributed to group membership with any better than 20% certainty. This is not particularly surprising given the small sample size of only six texts. Part of the difficulty with such an analysis is that a relatively limited sample prevents robust probability estimates for the subtle changes we would expect to find when analyzing

genre evolution. Unfortunately this is consistently problematic for language scholars because we often need to study a limited number of texts: limiting our sample size enables the possibility of thorough qualitative analyses of texts and is often necessary because of limited text availability (e.g. Shakespeare will not likely compose any more sonnets to the dark lady for contemporary scholars to study).

A statistical procedure known as "the bootstrap" (Efron, 1979; Efron et al., 2001) helps solve this difficulty. This procedure enables numerical estimates of the standard error of test statistics for a group. This supports better comparisons of groups by generating a large number of sample sets, each consisting of randomly drawn and reconstructed replacements from the original data set. The bootstrap uses a randomized selection of scores from the group to generate a new group of the same sample size (Efron et al., 2001). The bootstrap repeats this reformulation procedure many times (in this case 1,000 times) to create a large number of sample groups. From this large set of groups the test statistics are then drawn to make comparisons between the groups. One of the primary benefits of using the bootstrap procedure is that such comparisons may be made using any of the available test statistics (e.g., standard deviation, median, mode), rather than being limited to using the groups' mean as the standard of comparison as is the case with ANOVA.

To better analyze the changes in the "science fiction" genre between the corpora, I used "the bootstrap" procedure and made comparisons using the median and standard deviation scores for language dimension 2, "Writing for the Intellect." I chose the median as a better indicator of each

bootstrap group's central tendency because it is relatively insensitive to extreme scores in a distribution. In other words, the median is less affected than the mean by having one text with a markedly different score than the other texts in a group. This accounts for the differences seen between the mean and median scores in the second and third columns of table 6.1 (p. 207).

Using the bootstrap I also compared the standard deviations of language dimension 2 between the corpora. The standard deviation is an indicator of how consistently the authors of the texts used the language dimension. A comparison of standard deviations indicates whether or not the authors of the texts controlled their usage of the language dimension more or less carefully between the groups. That is, the standard deviation would tend to be smaller between the groups if the texts of the genre had become more similar to one another (i.e. the variance within the genre became smaller). On the other hand, if the standard deviation grew larger between the groups, this would indicate the texts of the genre had become more dissimilar in usage of the language dimension.

I used the bootstrap procedure to generate two distributions ($B = 1,000$ samples each) of median and standard deviation values, one distribution for the "science fiction" genre of each corpus. These large distributions provide for confidence intervals around the test statistics (median and standard deviation) and are shown in table 6.2. As you can see from these statistics, the "science fiction" authors of the Frown corpus tended to make more frequent usage of language dimension 2, "Writing for the Intellect." Furthermore, the Frown authors were more consistent

TABLE 6.2: Summary of the language dimension 2 bootstrap distribution scores for the "science fiction" genre of each corpus

	Median Distribution[1]			*SD* Distribution[1]		
Corpus	mean	med.	*SD*	mean	med.	*SD*
Brown	0.623	0.394	0.489	0.748	0.806	0.248
Frown	0.831	0.916	0.264	0.482	0.562	0.203

[1] $B = 1,000$ samples per bootstrap distribution

in their usage of it than the earlier Brown corpus science fiction authors. This indicates this particular genre underwent some changes between the corpora: the data suggests that the Frown corpus's "science fiction" authors[3] wrote texts more consistently aimed at priming intellectual reasoning processes in their audiences than did the authors of the earlier, more diffuse, Brown "science fiction" genre.

Such rhetorical understandings of genre change over time and between groups of texts are unique and potentially useful as we continue to explore relationships between the evolving notion of "genre" and the processes of rhetorical invention. Additionally, such analyses using whole texts as the base unit of analysis may enable application of rhetorical theory to new areas of interest, such as machine learning and text search and sort computer algorithms.

For example, let's compare two familiar genres commonly found in newspapers: press editorials and press reportage. I used the bootstrap

[3]Recall that the corpora in this study are both comprised of randomly selected text samples. The samples were collected with the aim of achieving a representative sample of *all* the texts published within that genre during the year.

TABLE 6.3: Summary of median bootstrap distribution scores for two genres of the Frown corpus

	Press Editorial			Press Reportage		
Dimension	mean	med.	*SD*	mean	med.	*SD*
1	-0.314	-0.327	0.069	-0.245	-0.257	0.099
2	0.478	0.440	0.144	-0.458	-0.504	0.109
3	-0.367	-0.321	0.198	0.549	0.588	0.110
4	-0.655	-0.686	0.193	-0.087	-0.123	0.205
5	0.258	0.265	0.085	0.254	0.248	0.120

$B = 1,000$ samples per bootstrap distribution

(Efron et al., 2001) to characterize the scores of these two genres from the Frown corpus[4] on the five language dimensions. The results, shown in table 6.3, indicate similarities between the authors' usage of two language dimensions (1 and 5) and differences in the usages of the other three dimensions.

Focusing our attention on language dimension 2, "Writing for the Intellect," we find the editorial writers prime intellectual processes more prevalently than do the authors of the newspaper reports. We can see this graphically in figure 6.2 (p. 212). This figure compares the bootstrap distribution of dimension 2 scores from the two genres.

Using this new understanding of one of the typical differences between these two genres, we can make predictive assessments about how individ-

[4]Prior to applying the bootstrap I randomly selected three "editorial" texts and five texts from the "reportage" genre to exclude from the procedure. These texts were reserved for use in subsequent regression analyses (not discussed as part of this research project). Text B25, used as an example later in this chapter, was among the randomly excluded texts.

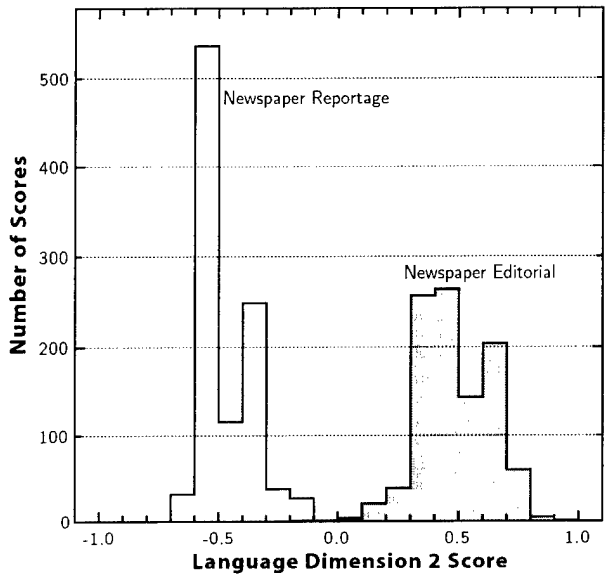


FIGURE 6.2: Bootstrap distribution comparison ($B = 1,000$) of median scores on language dimension 2, "Writing for the Intellect," for two genres of the Frown corpus

ual texts fit within the genre. For example, text B25 (excerpted from the *New York Times'* August 22, 1992, editorial page) has a score of 0.508 on language dimension 2. From this score one can calculate the likelihood the text would fall into each of the two distributions shown in figure 6.2. In this case, there is a 78.7% likelihood text B25 is an editorial (falling within the distribution of texts on the right side of figure 6.2) and very little likelihood ($\ll 0.00\%$) the text is of the "reportage" genre (the left-hand distribution of scores).

The predictive capabilities of such analyses could be further refined by using multivariate statistical processes such as logistic regression (Hosmer & Lemeshow, 1989; Kachigan, 1991). These processes would allow

the development of predictive formulæ that could more fully account for interactions among the multiple language dimensions within the genres and texts (see Collins et al., In Review, for an example use of this procedure).

What emerges from this type of analysis is a different understanding of texts and how individual texts fit and do not fit within particular "genres." This becomes an especially important consideration when such language scholarship is done in the context of rhetoric and composition pedagogy, a topic considered in the next section.

6.3 Altered view of rhetoric and composition

As should be clear from the preceding discussion and from other recent language scholarship (see chapter 2), computer analyses and statistical methods provide opportunities to make our understandings of rhetoric differently tangible. The software used in this analysis, for example, allows users to notice the rhetorical priming features in text by using color variation and frequency of occurrences, making the rhetorical theory available to the software user.[5] This book has attempted to demonstrate how access to this theory might be useful. However, such tangible understandings bring with them unintended consequences. It must remain an open question whether the benefits of such understandings outweigh the consequences of the altered focus they engender.

This is an especially important consideration as we think about using

[5]See appendix A for further discussion of the software.

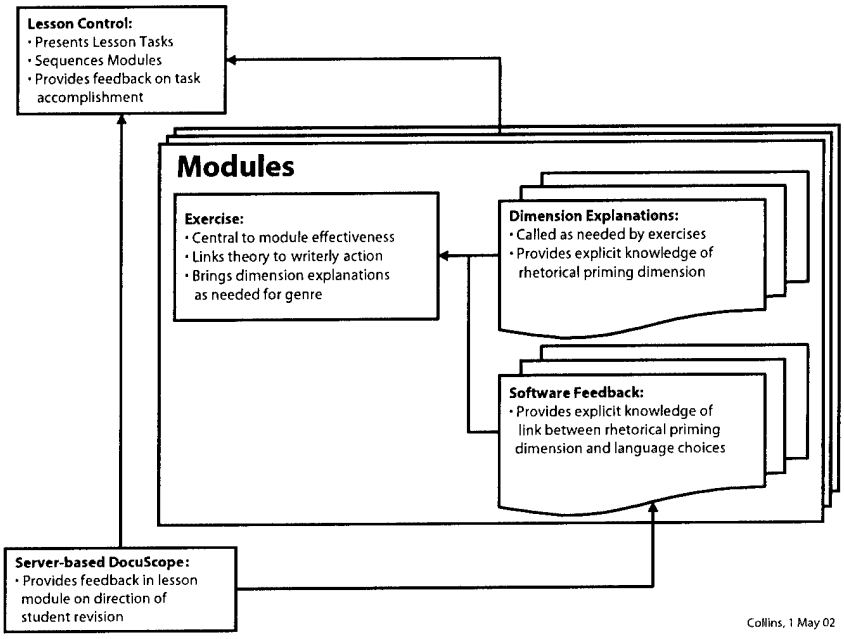


FIGURE 6.3: Example control diagram for "expert" coaching system. Lessons would be composed of sequenced modules designed to support student understanding and facilitate their control of the language dimensions (from Collins, 2002)

such understandings in writing pedagogy. For example, we could develop computer-based instructional systems that would help writing students learn to control the language dimensions this study has shown are important to the genres of professional writing (provided by figure 6.1). We could develop an instructional "domain model" (Wenger, 1987) surrounding writing, using the language dimensions as a framework for guiding the development and arrangement of lesson modules (figure 6.3) to support writers as they learn and practice the writing craft. We might integrate multi-media and the tangible feedback afforded by software (fig-

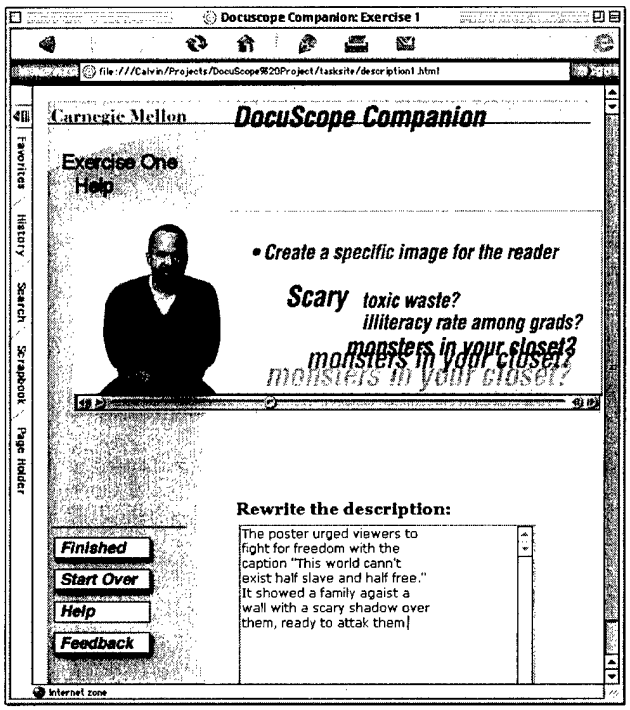


FIGURE 6.4: Example of web-based writing instruction interaction: the student has clicked on the "help" button, requesting information about responding to the writing prompt prior to requesting feedback from the computer (from Collins, 2002)

ure 6.4) to help guide students as they write texts in different genres and for different purposes. For example, the software could help guide students in writing editorials that are distinguishable from simple reports of events. Such systems would likely deepen student understandings of genre and could bring some limited benefits to the student (Cooper, 1999; Charney & Carlson, 1995; Hillocks, 1986, 1995). But at what consequence would these benefits come?

Such possibilities and questions represent the research challenge I

wish this book to offer those interested in the fields of rhetoric and composition. The study presented here shows tangible linkages between micro-level linguistic decisions and the macro-level workings of text. I have become convinced that control of these rhetorical linkages represents part of the challenge speakers and writers face as they design texts to meet their rhetorical situations. Like all design challenges, creating an appropriate, successful solution is a multivariate task that changes as the designer invents, produces draft text, and learns along the way. I believe there are better, yet-to-be invented ways to teach and support students in navigating these design tasks. Researching these ways and implementing them in beneficial ways is the challenge this study offers our academic community.

Appendices

A Description of DocuScope

DocuScope is ©Copyright 1998–2003 by Carnegie Mellon University. See page 219 for a list of the current development team. Parts of this description were written by Collins and Kaufer to accompany an analysis of *The Federalist* papers (Collins et al., In Review).

A.1 Frequently Asked Questions

What does DocuScope do?

DocuScope is designed to let people visualize and understand rhetorical priming in texts. It is *not* an attempt at artificial intelligence and the program does *not* "understand" or analyze anything it "reads." DocuScope simply goes through text documents and finds character strings that the creators of the software have indicated are relevant to rhetorical priming. DocuScope then displays these matches in ways that help users see the patterns in the texts. The software will also output quantifications of these patterns to comma-delimited text files for analysis in statistical packages. DocuScope is a Java application, running in the Java Runtime

Environments under Windows, Mac OS X, and on other Unix/Linux platforms.

What is rhetorical priming?

Rhetorical priming is a language theory developed to help writers predict effects of their texts on readers. We are at work on a book (Kaufer et al., In Press) that explains the language theory fully. The theory is described in some detail in chapter 4 (p. 80). Even without these references, we've observed students learning the language theory by using DocuScope and seeing how the computer tool "reads" and visualizes their texts.

Who is working on DocuScope?

DocuScope has been developed at Carnegie Mellon University's English Department to support writing courses taught to information design and professional writing students. The development team includes David Kaufer, Suguru Ishizaki, Brian Butler, Kerry Ishizaki, Milu Ritivoi, Pantelis Vlachos, and Jeff Collins.

Where can DocuScope be downloaded?

At this time it's not available to be downloaded. We have made DocuScope available free of charge to academic researchers interested in evaluating it. Since it is not a commercial product, we cannot guarantee support. If you are interested, contact David Kaufer (Kaufer@andrew.cmu.edu). DocuScope is owned by the authors and Carnegie Mellon University. You

will need to sign a release from Carnegie Mellon's technology transfer office before we can send you a copy of the software.

How can I learn to use DocuScope?

Over the four years of its classroom use, DocuScope has evolved toward becoming a mature text tagging and visualization system. It is beyond the scope of this document to describe the program's use in the writing classroom or in conducting rhetorical research on texts. A short tutorial is being tested and disseminated to students, educators and researchers learning to use the software. Contact Jeff Collins (jeff.collins@acm.org) for information about this tutorial.

We have given several workshops with DocuScope for writing teachers in the Pittsburgh public and independent schools. The response has been enthusiastic. DocuScope takes about an hour of face-to-face training to get teachers and students up to speed with it. We have not had the resources to provide true training and multi-institutional testing needed for large-scale phased dissemination.

A.2 Technical Information

Tagging Language Strings

DocuScope could be accurately described as flexible string-matching software: it can match strings up to any length. We designed the string-matcher to support human coding schemes. Thus, the string-matcher does no automatic analysis of its own, but rather implements whatever

coding of categories a human has supplied across a text, or corpus of texts. In our case, the coding categories have been based on our theory of rhetorical priming, as discussed in chapter 4.

The idea of a string-matcher that can match strings of different lengths is important because a series of words may not disambiguate themselves with respect to rhetorical action until deep into the series. Consider, for example, these two strings:

(50) On one hand there was a freckle. ("hand" refers to spatial relationship)

(51) On one hand there was evidence of fraud. ("hand" is part of a logical opening)

Rhetorically, the first sentence creates a spatial experience for a reader. The second creates a sense of engaging the reader's reasoning in an argument. These are quite different experiences for the reader, but the strings do not start to disambiguate until the sixth word.

To capture the difference, our string-matcher required an algorithm that holds in memory all strings that begin with "on one hand there was" while it explores the next word in the series, seeking a match that disambiguates the rhetorical function. When a disambiguation is found, the string can be isolated and assigned to a classification within a hierarchical classification scheme (section 4.5, p. 94).

The software we have built recognizes and tags contiguous strings that appear to do rhetorical work. In relation to theorists of written communication, our software captures an account of language that overlaps with

much of the language phenomena associated with writer-reader involvement, tying writers to readers through textual experience that can stand in for face-to-face experience (see chapter 2).

For several years now, the catalogs of strings have continued to grow incrementally in a cooperative process with students in our writing classrooms. While the rhetorical tagging program we designed has done the job we needed it to do, it has limitations. For example, it tags only contiguous language parts. It thus knows nothing about logical dependencies or about shifting speaking roles within textual dialog, for example.

Language Visualization

Differences in rhetorical priming at the string-level are often hard for the human eye to detect. Therefore, we attached our string-matcher to a visual interface that allows human coders to see its performance on actual texts.

In general, visualization systems attempt to tap people's natural strengths in rapid visual pattern recognition to support performance in activities that involve information processing (Gershon & Eick, 1997; Spoerri, 1993; Keim, 2002). As has been demonstrated in many areas, graphics are useful for enhancing human performance because complex cognitive tasks (e.g., comparing numbers from a table) may be replaced by perceptual inferences (e.g., perceiving the relative height of two adjacent bars).

One of the key considerations in designing visualization systems, as Lohse (1993) suggests, is to "facilitate and direct attention to visual feature(s) that communicate the requisite information when it is needed

during the task" (p. 385, emphasis added). While visualization technologies have been shown to support thinking and decision-making in many non-technical, multivariate areas such as management and legal studies (Card et al., 1999; Verheij, 1999; Suthers et al., 1997; Wright, 1997), they have not yet been widely employed in language study, outside of concept mapping and annotation or collaboration support (Brush et al., 2001; de Haan et al., 1999; Morris et al., 1999; Wolfe & Neuwirth, 2001).

There are several likely reasons for this, including the necessity of reading texts closely and serially (Chenoweth, 1997; Flower, 1988); the difficulty aggregating textual understandings visually (Wise et al., 1995; Neuwirth et al., 2001; Havre et al., 2002; Becker, 1997); and the difficulty in providing closely-coordinated and appropriate task representations and graphical support for reading, writing, and revising performance (Neuwirth et al., 1998; Ericsson & McGee, 2001).

The extraordinary modularity of texts requires people to manage their attention in support of their current writing and reading task(s) (Hayes, 1996; Wallace, 1991). While empirical observation of these processes is difficult, researchers have outlined several of the constructive reading and writing processes (Haas & Flower, 1988; Kintsch, 1989; Haas, 1996). Underlying these processes are four, interdependent cognitive tasks that we hypothesize could be supported with visualization technology: (1) fixation and (2) spatial/temporal attention that leads toward (3) explicit and implicit "seeing" and (4) conscious perception (Lohse, 1993; Chun & Wolfe, 2001).

To support our writing students and to begin to test our theories, we've

associated the different rhetorical priming effects with specific colors and, using colors, weighting, and positioning, have developed several visualization schemes within DocuScope to support some of the cognitive processes that underlie reading and writing. We have not formally evaluated the effectiveness of these schemes. In this paper we provide four screen captures to illustrate some of the visualization schemes that our students have found beneficial (figures A.1–A.4).

Software Operation

DocuScope has two integrated applications. One application is for single text visualization (STV) and another for multiple-text visualization (MTV). In writing classrooms, students tend to start in MTV to locate significant rhetorical trends across the classroom of texts (that is, they analyze multiple texts at once). Students then drill down into single texts to understand how individual student writers (and texts) produce these trends.

The following screen shots were captured while analyzing the Brown corpus (Francis & Kucera, 1979; Hofland et al., 1999). These shots illustrate four of the methods of text visualization offered by the software.

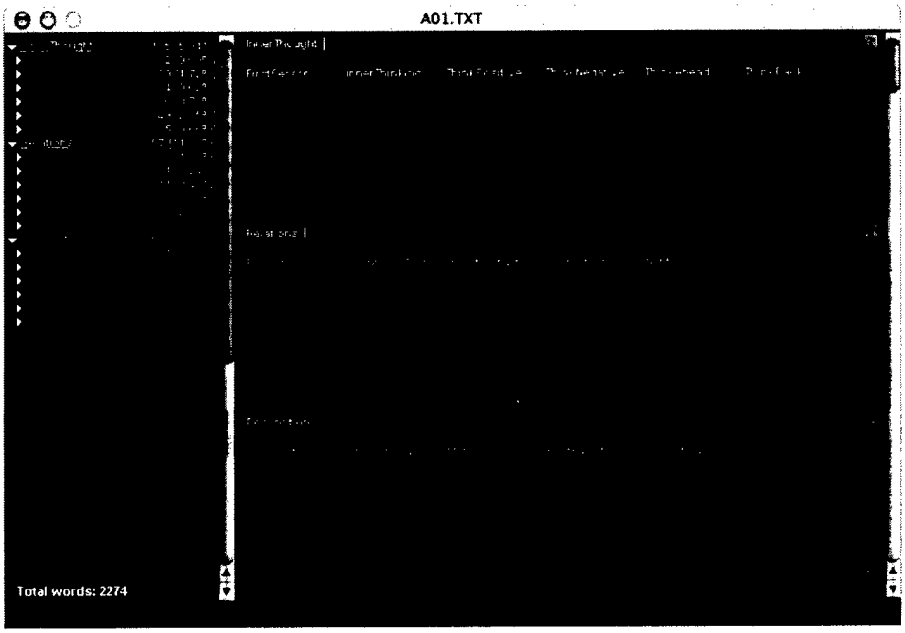


FIGURE A.1: Single Text Visualizer (STV): Dimensional View

The STV dimensional view shows a single text that has been tagged by the software. The text's score on each of the 18 variables is listed at the left of the screen. Each of the 18 matrices at the right of the screen is a view of the same portion of the text, but each highlights a different category of rhetorical priming.

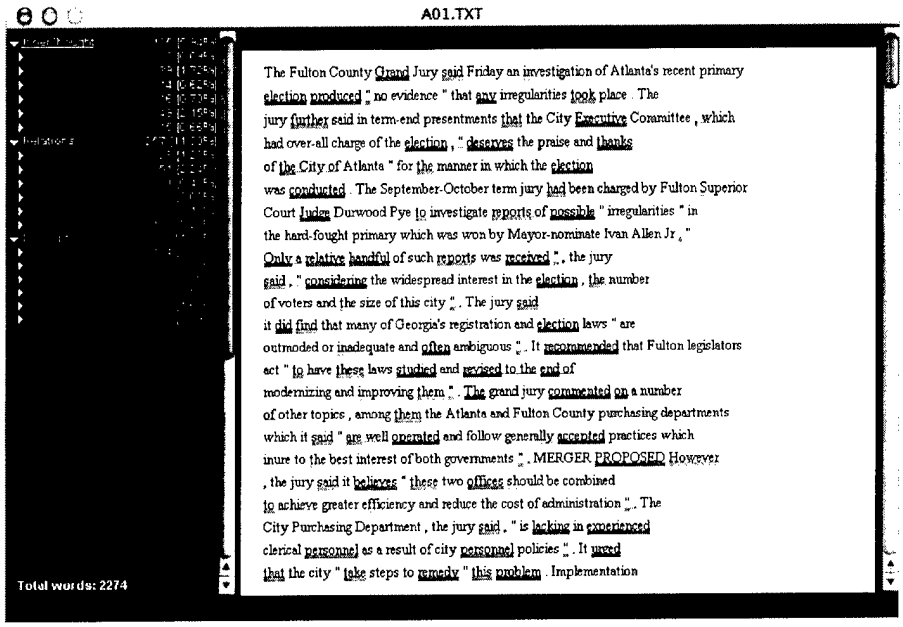


FIGURE A.2: Single Text Visualizer (STV): Text View

The STV text view shows a single text that has been tagged by the software. The text's score on each of the 18 variables is listed at the left of the screen. The text itself, with tagged strings underlined, appears at the right of the screen. The tags may be toggled off or on by clicking on the variable name in the list at left.

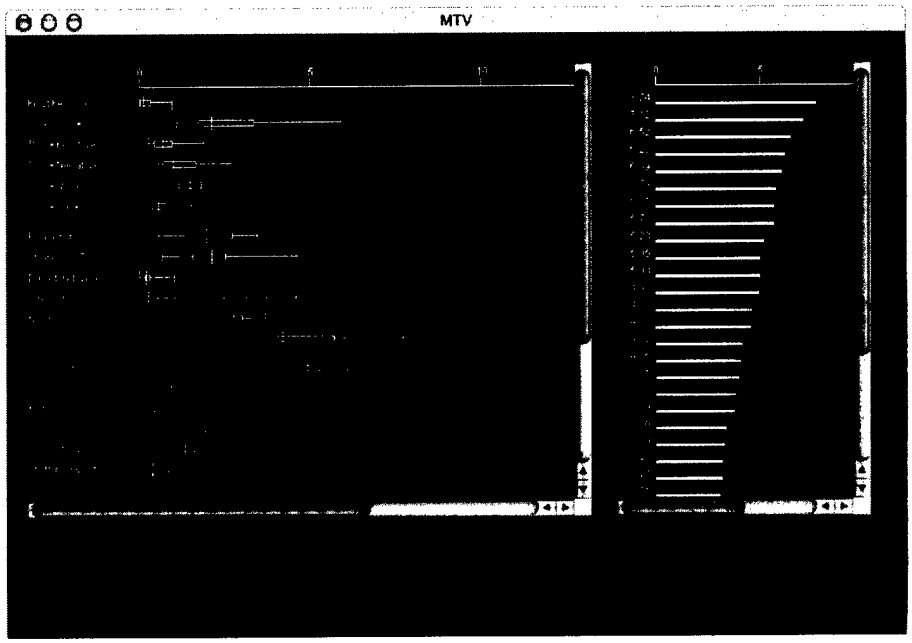


FIGURE A.3: Multi-Text Visualizer (MTV): Range View

The MTV range view summarizes the scores of all the tagged texts. The 18 priming categories are listed at the left of the screen with a modified boxplot* indicating the range of scores on each variable. The texts are listed at the right of the screen, sorted by the currently-highlighted variable ("LinearGuidance" in this shot, indicated by the different color of the variable name). The texts may be sorted into a priori groups; the software maintains these groups and assigns a different color to each group for easy separation, as indicated in the lower right-hand corner of the screen shot.

* The endpoints of these "modified" boxplots represent the high and low values, including outliers (normally the endpoints of boxplots represent the inner-quartile range). The "fences" are marked by the yellow indications above and below the boxplots.

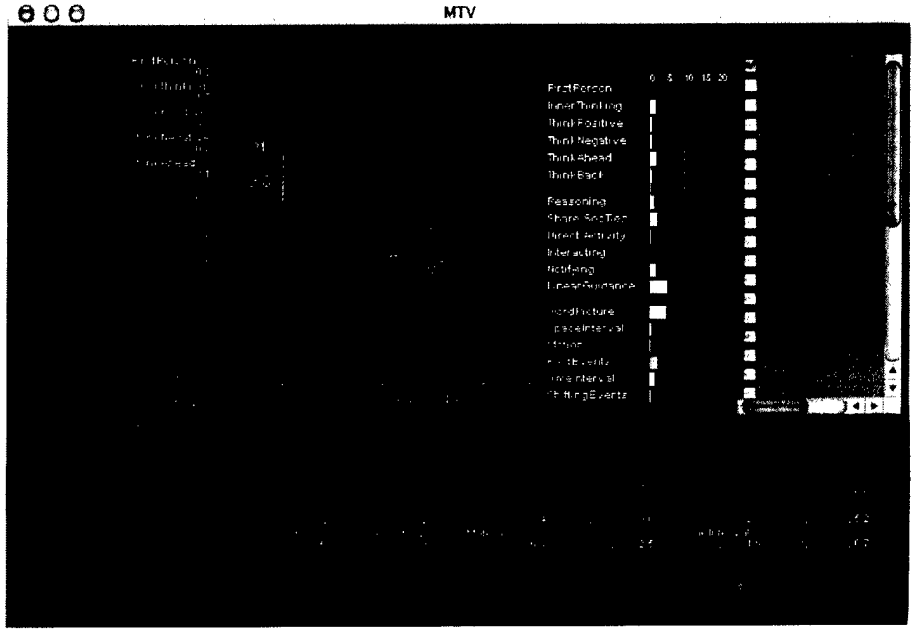


FIGURE A.4: Multi-Text Visualizer (MTV): Map View

The MTV map view shows all the texts in a collection of texts. The 18 response variables are listed along the X and Y axes. Clicking on a variable on an axis plots the texts' scores on that variable along that axis. The resulting plot may be scaled, as necessary. The texts are listed at the right of the screen. Clicking on an individual text highlights that text within the plot and generates the bar graph shown just to the left of the text names. The texts may be sorted into a priori groups and the software colors them for easy visual separation, as indicated in the lower right-hand corner of the screen shot.

B Putting the corpora texts into machine-readable form

This appendix describes the changes made to the corpora and provides the basic script used to break the electronic versions of the corpora from the CD-ROM into individual text files for analysis. Refer to section 3.1.1 for an explanation of how these processes fit into the overall methodology of the study.

I used widely-available electronic versions of the Brown and Frown corpora (Francis & Kucera, 1979; Hundt et al., 1999) from the 2nd edition of the International Computer Archive of Modern and Medieval English (ICAME) CD-ROM (Hofland et al., 1999).

B.1 Removing markup tags from the corpora

To facilitate various researcher interests, the CD-ROM includes three versions of the Brown corpus with varied markup systems to indicate some particular typographic and language usages. I used Bergen's text format

II, which has most of the typographical markup information removed but includes tags specifying some usages in the text.

For example, in the Brown corpus a tilde character (~) precedes acronyms, as in line 120 from Brown text A01:

> Robert Snodgrass, state ~GOP chairman, said a meeting...

Similarly, Roman numerals had been replaced with arabic equivalents and distinguished from other numerals with a back-slash (/) and comma markup system, as in line 126 from Brown text B20:

> the French Revolution the body of Henry /4,, who had...

I removed all such marks in order to put the texts into their original, text-only form (e.g. "Snodgrass, state GOP chairman..." and "the body of Henry IV, who had..."). Likewise, other markup symbols indicating headlines, font changes and other usages (see Hofland et al., 1999) were eliminated, preserving the original text as closely as possible. I also replaced the marked mathematical formulæ in the corpus with the string "fmula" to prevent the tagging software from incorrectly including portions of a formula (such as the variable a) as part of a priming string.

To make these changes to the texts on the CD-ROM I used BBEdit (Siegel et al., 2002), a Macintosh text editor, to make a series of searches and replacements. To ensure the accuracy of each replacement, I reviewed the summary of changes provided by the software prior to comitting the changes to the texts. This procedure was not perfect, missing some of the mis-typed tags in the corpus (e.g. places where the corpus coders had transposed a space and tag symbol, for example). I attempted to account

for miscodings by using additional search patterns. Tables B.1 and B.2 list the searches and replacements I used to modify the two corpora.

In addition to removing the information tags and formulæ in the Frown corpus, I also accepted marked-up corrections to misspellings and grammatical substitutions. For example, line 155 of Frown text A39 was originally published as "It's all to easy to dismiss. . . ." The electronic version of the corpus includes the likely correction, "It's all too easy to dismiss. . . ." I accepted all such tagged corrections to the Frown corpus because the version of the Brown corpus on the CD-ROM had had similar misspellings and substitutions corrected (but they were not marked in the texts).

I attempted to make no changes of my own to the corpora. That is, nothing was changed that had not been previously tagged in the electronic versions from the CD-ROM.

B.2 Separating corpora into constituent text files

To facilitate tagging and statistical analysis, I split each corpus into 500 individual text files. This was accomplished using a series of AppleScripts written for BBEdit (an example is shown below).

There was one notable problem with the Frown corpus: the lines from texts F22 and F23 had been merged onto the same lines in the original file. Fortunately, the line numbers for the texts were preserved, making it possible to fseparate and re-sort the lines of the file according to each

TABLE B.1: List of the search strings used for the Brown Corpus. The searches were performed in the sequence shown

Ref	Search Pattern[a]	Replace Pattern[b]
1	~	
2	**h	...
3	**f	fmula
4	&.	.
5	&	.
6	+	&
7	^	
8	\	
9	|	
10	@	
11	/1,	I
12	/2,	II
13	/3,	III
14	/4,	IV
15	/5,	V
16	/6,	VI
17	/8,	VIII
18	/12,	XII
19	_([^_]*)_	\1
20	#([^#]*)#	\1
21[c]	/[1-9]+,	Roman numerals
22	[0-9][\s]+-	
23[d,e]	([A-R][0-9][0-9])[\s]0010[\s]+1[\s]+	<#BROWN:\1\\>
24	[A-R][0-9][0-9]\s[0-9]+[\s]*[0-9]*[\s]+	
25[f]	<#BROWN:[^>]*>	

[a] Searches 1–18 are literal; 19–25 are grep
[b] Empty replacement values indicate deletions
[c] Hand-replacement of remaining Roman numerals is necessary
[d] Sets a string at the beginning of each Brown text to be used subsequently to break the corpus into individual text files
[e] A coding error in the corpus numbering prevents this search from properly setting a string for text L15
[f] Removes the string (set by search 23) from the beginning of each text

TABLE B.2: List of the search strings used for the Frown Corpus. The searches were performed in the sequence shown

Ref	Search Pattern[a]	Replace Pattern[b]
1	<[ph][/_|]*>	
2	<quote[_/|]*>	
3	<t[fp][_/|]*>	
4	<foreign[_/|]>	
5	<O[/_]>formula[e]?<O[/_]>	fmula
6	<O_>[^<]*<O/>	
7	<TranslitG[/_]>	
8	<s[bp][_/]>	
9	<\&[_/]>[^<]*<\&[_/]>	
10	<\&\\|>sic!	[sic]
11[d]	<sic!>	[sic]
12	<\?[/_]>	
13	<\}_><-\\|>[^<]*<\+\\|>([^<]*)<\}/>	\1
14	<*[/_]>([aeioucnzAEIOUCNZ])-[^<]*<*[_/]>	\1
15	<*_>([aeoAEO]*)-ligature<*/>	\1
16	<*_>[^<]*<*/>	
17[c]	<FROWN	<#FROWN
18[d]	<\&[^s]*sic[!]*[>]*	[sic]
19	<\{_><-\\|>[^<]*<\+\\|>([^<]*)<[\{\}]/>	\1
20	<[^>]*><-\\|>[^<]*<\+\\|>([^<]*)<\}/>	\1
21	<[^d]+degree[^>]*	
22[d]	<*|[/_]>([aeioucnzAEIOUCNZ])-[^<]*<[*]*|[_/]>	\1
23	it's <\}([^>]+)>	\1
24	<-\\|>to	
25	<\?>	

[a] All searches are grep

[b] Empty replacement values indicate deletions

[c] Standardizes tags at the beginning of each Frown text to be used subsequently to break the corpus into individual text files

[d] Searches 11, 18, and 22–28 correct for tagging errors in the corpus

TABLE B.2: Continued

Ref	Search Pattern	Replace Pattern
26[d]	<quote/	
27	<_*>a-circteau	ateau
28	<p/<	
29	<[^#]	fmulaLT
30[e]	[^\r]F23([\s]+)	\rF23\1
31	[A-R][0-9][0-9]\s[0-9]*[\s]*[0-9]*[\s]+	
32[f]	<#FR[^>]*>	

[d] Searches 11, 18, and 22–28 correct for tagging errors in the corpus

[e] Lines of texts F22–23 are intermixed. Search 30 inserts a linebreak so the lines may be properly separated and sorted

[f] Removes the string from the beginning of each text

text's line numbers. The grep searches necessary for making this correction are included in the list of changes in table B.2.

Example AppleScript used to separate corpus texts

```
tell application "BBEdit"
    activate
    set textNum to 1
     repeat until textNum = 44 --reset number for each genre
          find "<#BROWN:[^>]*>[^<]*" searching in text 1 of text
               window 1 options {search mode:grep, starting at top:false,
               case sensitive:false} with selecting match
          copy selection
          make new text window with properties {contents:current clipboard}
          set textPath to "Calvin:Corpora:Brown_broken:" --file path
          set textName to textPath & "A" & textNum & ".txt" --file name
          save text window 1 to file textName
          <<event ascrgdte>> 0
          close text window 1
          set textNum to 1 + textNum
     end repeat
end tell
```

C Variable Scores by Genre

The following tables summarize the results of the counting procedure applied to the Brown corpus. See section 3.1.3 (p. 40) for more information about the procedure and for an explanation of the statistics provided in the following tables.

TABLE C.1: *First Person* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.2166	0.130	0.2395	2.459	<0.001
B	27	0.3700	0.130	0.4505	2.025	<0.001
C	17	0.1482	0.080	0.1940	1.466	0.001
D	17	0.5880	0.260	0.7750	1.415	0.001
E	36	0.2394	0.000	0.4992	6.624	<0.001
F	48	0.2815	0.085	0.4883	5.772	<0.001
G	75	0.6179	0.350	0.8541	6.749	<0.001
H	30	0.1810	0.000	0.3932	5.150	<0.001
J	80	0.1477	0.000	0.3475	14.058	<0.001
K	29	0.9370	0.430	1.1590	3.315	<0.001
L	24	1.2220	0.745	1.2500	2.023	<0.001
M	6	0.9170	0.960	0.3500	0.310	0.431
N	29	1.3470	0.830	1.4600	2.341	<0.001
P	29	1.7310	0.870	1.6930	2.413	<0.001
R	9	1.4470	1.470	0.9410	0.378	0.326

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.2: *Inner Thinking* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.8750	1.800	0.7040	0.656	0.081
B	27	2.9510	2.830	0.6020	0.394	0.351
C	17	2.8440	2.790	0.4880	0.191	0.882
D	17	3.3200	3.090	0.8710	1.410	0.001
E	36	2.4780	2.485	0.8430	0.133	0.978
F	48	2.6490	2.470	1.1260	0.851	0.027
G	75	2.8899	2.830	0.7549	0.723	0.057
H	30	2.0140	2.110	0.8420	0.338	0.480
J	80	2.8910	2.780	0.9030	0.647	0.088
K	29	2.5600	2.400	0.6220	0.314	0.527
L	24	3.0150	2.940	0.8350	0.399	0.338
M	6	3.0300	2.575	0.9800	0.537	0.096
N	29	2.6486	2.500	0.5375	0.650	0.081
P	29	2.9620	2.960	0.6330	0.287	0.598
R	9	2.9240	2.900	0.5280	0.445	0.215

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.3: *Think Positive* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.6157	0.565	0.2974	0.735	0.051
B	27	0.7978	0.740	0.2709	0.628	0.091
C	17	1.4360	1.340	0.4530	0.224	0.791
D	17	0.7894	0.780	0.4042	0.446	0.248
E	36	0.8853	0.730	0.5022	0.585	0.118
F	48	0.8213	0.785	0.3688	0.387	0.376
G	75	0.8492	0.760	0.3586	1.112	0.006
H	30	0.6110	0.505	0.3377	0.567	0.129
J	80	0.5718	0.435	0.3942	2.152	<0.001
K	29	0.7579	0.700	0.3440	0.767	0.041
L	24	0.5279	0.495	0.2762	0.694	0.061
M	6	1.1500	1.065	0.3480	0.265	0.544
N	29	0.6424	0.610	0.3130	0.476	0.221
P	29	0.9403	0.870	0.3756	1.622	<0.001
R	9	1.0460	0.860	0.5380	0.362	0.358

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.4: *Think Negative* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.9684	0.860	0.5044	0.674	0.073
B	27	1.6110	1.610	0.5250	0.233	0.777
C	17	1.2610	1.050	0.5410	0.706	0.053
D	17	1.5010	1.270	0.7910	0.681	0.062
E	36	0.6544	0.560	0.4053	1.915	<0.001
F	48	1.3680	1.260	0.6950	0.823	0.031
G	75	1.4969	1.460	0.6331	0.223	0.820
H	30	0.5910	0.565	0.4005	0.522	0.169
J	80	1.0904	0.940	0.7591	3.230	<0.001
K	29	1.4200	1.410	0.5630	0.162	0.939
L	24	1.3496	1.365	0.3701	0.300	0.555
M	6	1.3350	1.285	0.2850	0.420	0.211
N	29	1.3934	1.280	0.5058	0.455	0.250
P	29	1.4531	1.300	0.5042	1.210	0.003
R	9	1.6370	1.680	0.5600	0.513	0.140

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.5: *Think Ahead* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.4355	1.330	0.4704	0.745	0.049
B	27	1.6144	1.550	0.4007	0.511	0.179
C	17	0.8229	0.850	0.2820	0.279	0.602
D	17	1.1518	1.120	0.2826	0.406	0.312
E	36	1.2597	1.200	0.5378	0.252	0.720
F	48	1.2423	1.125	0.4755	0.777	0.041
G	75	1.1275	1.100	0.3718	0.515	0.186
H	30	1.6050	1.330	0.8200	0.827	0.029
J	80	1.0828	1.030	0.4560	0.300	0.575
K	29	1.0928	1.000	0.3972	0.876	0.022
L	24	1.1933	1.215	0.3063	0.528	0.597
M	6	1.1850	1.235	0.2460	0.248	0.599
N	29	1.0000	1.020	0.2570	0.534	0.157
P	29	1.1017	1.150	0.2708	0.262	0.678
R	9	0.9360	0.900	0.3050	0.297	0.514

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.6: *Think Back* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.5432	0.505	0.2482	0.632	0.093
B	27	0.6233	0.580	0.3028	0.403	0.334
C	17	0.4994	0.430	0.2044	0.371	0.382
D	17	0.4729	0.510	0.2431	0.235	0.754
E	36	0.3481	0.310	0.2358	0.519	0.175
F	48	0.5725	0.525	0.3555	1.444	0.001
G	75	0.6263	0.560	0.3705	1.828	<0.001
H	30	0.3580	0.270	0.2582	1.087	0.006
J	80	0.4379	0.395	0.2215	1.440	0.001
K	29	0.9128	0.890	0.3471	0.362	0.421
L	24	0.8833	0.825	0.4354	0.585	0.115
M	6	0.9050	0.935	0.2311	0.511	0.115
N	29	0.7072	0.680	0.2558	0.337	0.481
P	29	0.8159	0.870	0.3546	0.533	0.158
R	9	0.6120	0.450	0.3870	0.445	0.215

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.7: *Reasoning* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.7740	1.680	0.7310	0.589	0.117
B	27	2.7860	2.640	0.5870	0.287	0.595
C	17	2.5150	2.600	0.4740	0.344	0.445
D	17	3.0920	2.940	0.8420	0.714	0.051
E	36	2.4000	2.270	0.9410	0.276	0.637
F	48	2.4920	2.535	0.7510	0.253	0.722
G	75	2.8253	2.630	0.7637	1.906	<0.001
H	30	1.9970	1.900	0.7920	0.252	0.716
J	80	2.7869	2.770	0.8783	0.194	0.891
K	29	2.5050	2.360	0.7660	0.436	0.279
L	24	2.6170	2.350	0.7200	0.864	0.022
M	6	3.5050	3.430	0.6190	0.494	0.126
N	29	2.2510	2.250	0.6120	0.335	0.487
P	29	2.9180	2.810	0.7130	0.374	0.393
R	9	2.7670	2.800	0.5000	0.266	0.595

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.8: *Share Social Ties* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.8390	1.870	0.7530	0.684	0.069
B	27	2.6760	2.540	0.6860	0.325	0.508
C	17	1.9640	1.820	0.6760	0.980	0.010
D	17	3.4120	3.470	0.9760	0.218	0.809
E	36	1.8720	1.845	0.7490	0.146	0.964
F	48	2.0440	1.865	0.9270	0.902	0.020
G	75	2.4930	2.280	0.9840	0.744	0.050
H	30	2.9690	2.600	1.3190	0.871	0.022
J	80	2.1590	2.020	1.0120	0.848	0.028
K	29	1.3352	1.250	0.4355	0.548	0.145
L	24	1.1463	1.145	0.3477	0.275	0.630
M	6	1.4880	1.475	0.5450	0.157	0.905
N	29	1.0890	1.010	0.4127	0.777	0.038
P	29	1.3624	1.360	0.4116	0.451	0.256
R	9	1.3980	1.370	0.3690	0.165	0.910

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.9: *Direct Activity* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.2345	0.210	0.1489	0.864	0.024
B	27	0.2707	0.300	0.1156	0.555	0.138
C	17	0.1706	0.170	0.0964	0.446	0.248
D	17	0.2724	0.250	0.2027	0.478	0.205
E	36	0.6260	0.375	0.6510	2.821	<0.001
F	48	0.2598	0.215	0.1979	2.141	<0.001
G	75	0.2176	0.180	0.1486	2.204	<0.001
H	30	0.4227	0.245	0.3873	1.985	<0.001
J	80	0.2971	0.250	0.2229	2.423	<0.001
K	29	0.2272	0.180	0.1500	0.755	0.044
L	24	0.2929	0.260	0.1833	0.452	0.250
M	6	0.2617	0.230	0.1146	0.268	0.536
N	29	0.1941	0.130	0.1351	0.773	0.040
P	29	0.2372	0.240	0.1064	0.290	0.588
R	9	0.1622	0.160	0.1063	0.244	0.672

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.10: *Interacting* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.1964	0.170	0.1527	0.834	0.029
B	27	0.5989	0.520	0.3923	0.429	0.288
C	17	0.2888	0.250	0.2328	0.774	0.350
D	17	0.7420	0.270	0.7860	1.968	<0.001
E	36	1.1240	0.725	1.5010	2.984	<0.001
F	48	0.4630	0.270	0.7230	6.837	<0.001
G	75	0.4368	0.310	0.4400	5.617	<0.001
H	30	0.2840	0.105	0.6070	5.293	<0.001
J	80	0.1982	0.090	0.2802	7.422	<0.001
K	29	0.9248	0.760	0.4735	0.639	0.086
L	24	1.4220	1.380	0.7050	0.193	0.884
M	6	1.3420	1.320	0.6100	0.169	0.882
N	29	1.1070	1.080	0.5510	0.678	0.068
P	29	1.4840	1.330	0.8100	0.671	0.071
R	9	1.2280	1.270	0.8720	0.342	0.403

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.11: *Notifying* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	2.5566	2.505	0.4904	0.362	0.429
B	27	2.8352	2.750	0.4243	0.226	0.798
C	17	2.7529	2.750	0.3388	0.187	0.890
D	17	2.9110	2.790	0.5790	0.367	0.391
E	36	2.7522	2.765	0.5490	0.321	0.517
F	48	2.9360	2.975	0.6557	0.457	0.255
G	75	2.8324	2.830	0.5825	0.250	0.737
H	30	2.8367	2.860	0.4970	0.182	0.904
J	80	3.1839	3.175	0.6634	0.569	0.136
K	29	2.0603	2.000	0.3793	0.317	0.521
L	24	2.1783	2.080	0.3890	0.650	0.079
M	6	1.9800	1.855	0.3560	0.387	0.262
N	29	1.9652	1.840	0.3629	0.463	0.239
P	29	1.9624	1.870	0.4574	0.326	0.507
R	9	2.3140	2.260	0.3770	0.536	0.120

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.12: *Linear Guidance* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	3.9350	3.730	0.9810	0.761	0.044
B	27	3.8650	3.780	0.9170	1.155	0.004
C	17	4.2280	4.150	0.8980	0.387	0.349
D	17	4.7830	4.760	1.3350	0.478	0.205
E	36	2.9080	2.365	1.4050	1.162	0.004
F	48	4.3390	4.050	1.5840	0.578	0.126
G	75	4.9800	5.030	1.2590	0.374	0.407
H	30	3.0510	2.915	1.0450	0.452	0.255
J	80	3.2120	3.120	1.2280	1.070	0.008
K	29	7.0000	7.060	1.6460	0.265	0.668
L	24	6.9850	6.925	1.3220	0.249	0.720
M	6	7.2380	7.170	0.6360	0.256	0.573
N	29	7.7870	7.700	1.6640	0.225	0.803
P	29	7.6780	7.680	1.5460	0.322	0.513
R	9	6.1570	4.860	2.3640	0.431	0.235

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.13: *Word Picture* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	5.2790	5.130	1.5000	0.344	0.472
B	27	4.8040	4.560	1.4370	0.577	0.120
C	17	5.5490	5.620	1.1690	0.212	0.826
D	17	3.9060	3.820	1.4210	0.267	0.641
E	36	6.2250	6.250	2.4980	0.706	0.060
F	48	6.4490	6.495	2.3730	0.716	0.058
G	75	4.6000	4.330	1.7780	1.387	0.001
H	30	3.7650	3.395	1.9650	1.148	0.004
J	80	3.8630	3.575	1.8940	0.786	0.040
K	29	7.3690	7.580	1.7780	0.274	0.637
L	24	6.3210	6.650	1.5560	0.660	0.075
M	6	4.7080	4.780	1.2740	0.300	0.456
N	29	7.5560	7.420	2.0250	0.288	0.594
P	29	6.6000	6.050	1.9260	0.658	0.077
R	9	5.4720	5.470	1.4080	0.138	0.959

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.14: *Space Interval* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.1466	1.080	0.5446	1.711	<0.001
B	27	0.9600	0.920	0.3521	0.749	0.045
C	17	0.9841	0.910	0.3570	0.492	0.189
D	17	0.9760	0.870	0.4760	1.621	<0.001
E	36	1.4720	1.315	0.7580	0.971	0.013
F	48	1.2671	1.120	0.6863	1.765	<0.001
G	75	1.0228	0.870	0.5746	4.116	<0.001
H	30	0.7790	0.695	0.4982	1.044	0.008
J	80	0.9941	0.880	0.5909	2.388	<0.001
K	29	2.2260	2.280	0.5910	0.303	0.552
L	24	2.2380	2.165	0.9310	0.600	0.106
M	6	1.3320	1.165	0.4180	0.411	0.224
N	29	2.4910	2.240	0.6780	1.003	0.010
P	29	1.7470	1.700	0.6190	0.274	0.638
R	9	1.3970	1.350	0.5850	0.406	0.275

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.15: *Motion* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.4118	0.370	0.2260	0.863	0.025
B	27	0.3863	0.340	0.1899	0.686	0.065
C	17	0.7753	0.700	0.3689	0.992	0.010
D	17	0.2706	0.300	0.1430	0.217	0.812
E	36	0.6611	0.440	0.4938	1.418	0.001
F	48	0.5127	0.470	0.3168	1.012	0.010
G	75	0.3327	0.260	0.2528	4.115	<0.001
H	30	0.2180	0.175	0.2156	1.275	0.002
J	80	0.3136	0.215	0.3508	6.425	<0.001
K	29	0.7507	0.690	0.3531	0.521	0.120
L	24	0.8004	0.720	0.4005	1.179	0.004
M	6	0.4933	0.480	0.1838	0.348	0.339
N	29	1.0066	0.840	0.4449	0.898	0.019
P	29	0.6907	0.580	0.4115	1.299	0.002
R	9	0.6467	0.590	0.2869	0.595	0.084

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.16: *Past Events* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	2.0740	2.075	0.7620	0.311	0.541
B	27	1.4400	1.420	0.4675	0.296	0.569
C	17	1.4980	1.550	0.4760	0.182	0.897
D	17	1.6120	1.360	0.6460	0.391	0.342
E	36	1.4990	1.425	0.7240	0.911	0.018
F	48	2.0690	1.880	1.0000	0.591	0.119
G	75	1.9740	1.740	0.9150	1.808	<0.001
H	30	1.4800	1.315	0.5660	1.126	0.005
J	80	1.7070	1.570	0.8339	2.001	<0.001
K	29	3.4920	3.380	0.7540	0.698	0.061
L	24	3.2720	3.135	0.9600	0.657	0.076
M	6	2.7550	2.695	0.5720	0.327	0.385
N	29	3.6980	3.840	0.8580	0.336	0.483
P	29	3.0520	3.030	0.7680	0.167	0.930
R	9	2.4340	2.110	0.9630	0.216	0.777

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.17: *Time Interval* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	1.4077	1.370	0.3332	0.269	0.665
B	27	1.2611	1.260	0.2326	0.332	0.494
C	17	1.3900	1.400	0.3012	0.529	0.151
D	17	1.2650	1.100	0.5330	0.505	0.175
E	36	1.4153	1.225	0.5837	1.207	0.003
F	48	1.3748	1.280	0.4897	1.083	0.007
G	75	1.3288	1.260	0.3973	1.276	0.002
H	30	1.6680	1.515	0.6760	0.456	0.248
J	80	1.5216	1.400	0.8381	1.830	<0.001
K	29	1.0779	1.050	0.2916	0.561	0.134
L	24	1.0842	1.070	0.3209	0.135	0.975
M	6	1.0733	1.105	0.1297	0.320	0.403
N	29	0.9762	0.920	0.2953	0.312	0.532
P	29	0.9897	1.010	0.2313	0.239	0.757
R	9	1.2011	1.250	0.2850	0.199	0.832

[a] 95% Confidence level for the A–D normality test statistics

TABLE C.18: *Shifting Events* summary statistics for the genres of the Brown corpus

					Anderson–Darling[a]	
Genre	n	Mean	Median	SD	A^2	p
A	44	0.8852	0.845	0.2985	0.887	0.021
B	27	0.8033	0.800	0.2119	0.139	0.971
C	17	0.7212	0.700	0.2008	0.309	0.522
D	17	0.6647	0.650	0.2271	0.311	0.519
E	36	0.7458	0.645	0.2729	1.641	<0.001
F	48	0.8990	0.900	0.3263	0.859	0.025
G	75	0.8048	0.760	0.3132	0.706	0.063
H	30	0.6613	0.620	0.3268	1.034	0.009
J	80	0.5964	0.540	0.2703	1.518	0.001
K	29	1.0148	1.040	0.2757	0.430	0.288
L	24	1.1192	1.130	0.3126	0.250	0.716
M	6	0.8417	0.735	0.2342	0.661	0.042
N	29	1.0507	1.070	0.2460	0.224	0.805
P	29	0.9638	0.910	0.3306	0.301	0.557
R	9	1.0289	1.050	0.2392	0.223	0.753

[a] 95% Confidence level for the A–D normality test statistics

D Factor Scores by Genre

Table D.2 provides summary statistics for each of the five extracted factors by group of texts (Brown half a; Brown half b; and Frown texts). For convenient reference I've listed of genres in the corpora in table D.1 below.

TABLE D.1: List of the genres and the number of texts (n) per genre per group[a] in the corpora

Reference: Text Genre	n_a	n_b	n_f
A: Press Reportage	22	22	44
B: Press Editorials	14	13	27
C: Press Reviews	9	8	17
D: Religion	9	8	17
E: Skills and Hobbies	18	18	36
F: Popular Lore	24	24	48
G: Belles Lettres, Biography, Memoirs	38	37	75
H: Miscellaneous	15	15	30
J: Learned	40	40	80
K: General Fiction	15	14	29
L: Mystery and Detective Fiction	12	12	24
M: Science Fiction	3	3	6
N: Adventure and Western	15	14	29
P: Romance and Love Story	15	14	29
R: Humor	5	4	9

[a] The group n is indicated by subscripts with a and b indicating the two randomly-drawn halves of the Brown corpus and f indicating the Frown corpus

TABLE D.2: Mean factor scores (and standard deviations) for the text genres of the corpora

Genre[a]	Factor 1 mean (*SD*)	Factor 2 mean (*SD*)	Factor 3 mean (*SD*)	Factor 4 mean (*SD*)	Factor 5 mean (*SD*)
A_a	-0.299 (0.67)	-1.052 (0.60)	0.243 (0.69)	0.008 (0.70)	0.279 (0.61)
A_b	-0.229 (0.45)	-0.889 (0.97)	0.053 (0.75)	-0.216 (0.71)	0.428 (0.57)
A_f	-0.234 (0.50)	-0.434 (0.57)	0.446 (0.73)	-0.240 (0.74)	0.239 (0.60)
B_a	-0.595 (0.40)	0.263 (0.63)	0.025 (0.60)	-0.616 (0.57)	0.065 (0.66)
B_b	-0.275 (0.38)	0.520 (0.76)	-0.269 (0.59)	-0.306 (0.74)	0.033 (0.76)
B_f	-0.286 (0.40)	0.505 (0.62)	-0.278 (0.61)	-0.793 (0.81)	0.232 (0.50)
C_a	-0.148 (0.25)	0.216 (0.81)	-0.861 (0.48)	0.766 (0.49)	-0.720 (0.92)
C_b	0.151 (0.65)	-0.251 (0.49)	-0.947 (0.66)	1.075 (0.51)	-0.739 (0.57)
C_f	0.285 (0.44)	-0.130 (0.53)	-0.740 (0.68)	0.819 (0.57)	-0.358 (0.66)
D_a	-0.620 (0.55)	1.266 (1.41)	-0.430 (0.54)	-0.165 (0.69)	-0.392 (1.49)
D_b	-0.330 (0.58)	0.423 (0.54)	-0.468 (0.58)	0.430 (0.49)	0.235 (0.66)
D_f	-0.657 (0.35)	0.528 (0.86)	-0.672 (0.64)	0.211 (0.93)	0.197 (0.94)
E_a	0.547 (1.16)	-0.638 (1.07)	-1.103 (0.79)	-0.933 (2.10)	-0.128 (1.23)
E_b	0.237 (0.93)	-0.512 (0.72)	-0.839 (0.70)	-0.930 (1.31)	-0.471 (0.98)
E_f	0.471 (0.96)	-0.680 (0.85)	-0.614 (0.93)	-0.400 (1.52)	-0.375 (0.94)
F_a	0.012 (0.89)	-0.202 (0.94)	-0.134 (1.09)	0.036 (1.02)	0.059 (0.87)
F_b	-0.062 (0.58)	-0.033 (1.01)	-0.102 (1.01)	0.099 (0.98)	0.003 (0.84)
F_f	-0.202 (0.69)	0.071 (0.83)	-0.474 (0.65)	0.049 (0.97)	0.088 (0.83)
G_a	-0.358 (0.65)	0.444 (0.79)	0.039 (0.96)	0.393 (0.70)	-0.041 (0.84)
G_b	-0.463 (0.57)	0.292 (0.91)	-0.019 (0.95)	0.292 (0.75)	-0.108 (0.85)
G_f	-0.393 (0.59)	0.069 (0.77)	0.121 (1.01)	0.479 (0.61)	-0.246 (0.87)
H_a	-0.870 (0.47)	-0.947 (0.86)	-0.301 (0.54)	-0.386 (0.76)	0.483 (0.59)
H_b	-0.805 (0.92)	-0.944 (0.97)	-0.549 (0.34)	-1.060 (2.06)	0.162 (0.73)
H_f	-0.821 (0.72)	-0.735 (1.25)	-0.368 (0.76)	-0.615 (1.31)	0.722 (0.66)

[a] Subscripts indicate from which group of texts statistics are drawn with a and b indicating the two randomly-drawn halves of the Brown corpus and f indicating the Frown corpus.

TABLE D.2: Continued

Genre[a]	Factor 1 mean (*SD*)	Factor 2 mean (*SD*)	Factor 3 mean (*SD*)	Factor 4 mean (*SD*)	Factor 5 mean (*SD*)
J_a	-0.561 (0.84)	-0.098 (1.02)	-0.561 (0.80)	0.215 (1.05)	0.589 (0.78)
J_b	-0.475 (0.76)	0.123 (0.90)	-0.524 (0.74)	0.345 (0.73)	0.682 (0.68)
J_f	-0.596 (0.68)	0.007 (0.94)	-0.596 (0.83)	0.330 (0.79)	0.728 (0.71)
K_a	0.693 (0.71)	0.063 (0.72)	1.017 (0.90)	0.081 (0.75)	-0.318 (1.26)
K_b	1.142 (0.77)	0.211 (0.62)	1.080 (0.64)	0.072 (0.54)	0.254 (0.95)
K_f	1.125 (0.83)	0.083 (0.78)	0.528 (1.00)	0.073 (0.80)	-0.786 (1.07)
L_a	0.882 (0.79)	0.602 (0.70)	1.111 (1.06)	-0.433 (0.47)	-0.130 (1.09)
L_b	0.954 (1.00)	0.102 (0.80)	1.227 (0.81)	-0.356 (0.51)	0.019 (1.48)
L_f	0.771 (0.64)	0.647 (0.75)	1.090 (0.83)	-0.080 (0.52)	-0.002 (0.77)
M_a	-0.080 (0.32)	0.992 (1.10)	0.829 (0.15)	-0.026 (0.44)	-0.430 (0.62)
M_b	0.397 (0.94)	0.692 (0.81)	0.943 (0.97)	-0.159 (0.15)	-0.922 (0.39)
M_f	0.520 (0.88)	0.694 (0.58)	1.146 (0.74)	-0.234 (0.31)	-0.284 (1.09)
N_a	1.695 (0.72)	0.172 (0.52)	0.678 (0.70)	0.347 (0.45)	0.075 (0.98)
N_b	1.148 (0.59)	0.160 (0.65)	1.036 (0.54)	0.128 (0.49)	-0.442 (1.30)
N_f	1.230 (0.67)	0.087 (0.49)	0.661 (0.49)	0.035 (0.60)	-0.233 (0.57)
P_a	0.935 (0.95)	0.578 (0.44)	0.802 (0.48)	-0.194 (0.59)	-0.815 (1.13)
P_b	0.582 (0.64)	0.639 (0.58)	0.902 (0.90)	0.215 (0.57)	-1.041 (1.38)
P_f	0.846 (0.77)	0.293 (0.88)	1.055 (1.21)	-0.438 (0.86)	-0.985 (1.22)
R_a	0.120 (0.26)	0.263 (0.28)	0.240 (1.02)	0.603 (0.88)	-1.339 (0.97)
R_b	0.661 (0.57)	0.594 (0.19)	0.324 (1.16)	0.321 (0.53)	-0.844 (1.11)
R_f	0.480 (0.98)	0.327 (0.59)	0.917 (1.09)	-0.503 (1.32)	-0.711 (1.28)

[a] Subscripts indicate from which group of texts statistics are drawn with a and b indicating the two randomly-drawn halves of the Brown corpus and f indicating the Frown corpus.

E Language Dimensions by Genre

The tables below provide the summary statistics for each language dimension by text genre. These statistics are represented graphically by the boxplots in chapter 5 (p. 127). The results shown in these tables were also subjected to the Tukey pairwise comparison test (Tukey, 1949). Refer to section 3.3.2 (p. 75) for a description and the results of this statistical test.

TABLE E.1: Summary stats for dimension 1, "Writing for the Eyes vs. Informing," scores by genre

		Brown Corpus		Frown Corpus	
Genre	n	Mean	SD	Mean	SD
A	44	-0.261	0.554	-0.234	0.495
B	27	-0.439	0.402	-0.286	0.398
C	17	-0.011	0.460	0.285	0.439
D	17	-0.470	0.553	-0.657	0.345
E	36	0.387	1.072	0.471	0.963
F	48	-0.031	0.741	-0.202	0.694
G	75	-0.411	0.613	-0.393	0.586
H	30	-0.848	0.708	-0.821	0.717
J	80	-0.509	0.806	-0.596	0.682
K	29	0.899	0.734	1.125	0.833
L	24	0.912	0.867	0.771	0.641
M	6	0.153	0.656	0.520	0.877
N	29	1.434	0.741	1.230	0.673
P	29	0.769	0.839	0.846	0.770
R	9	0.365	0.443	0.480	0.980

TABLE E.2: Summary statistics for dimension 2, "Writing for the Intellect," scores by genre

		Brown Corpus		Frown Corpus	
Genre	n	Mean	SD	Mean	SD
A	44	-0.964	0.808	-0.434	0.567
B	27	0.388	0.699	0.505	0.615
C	17	-0.007	0.688	-0.130	0.532
D	17	0.878	1.112	0.528	0.855
E	36	-0.579	0.906	-0.680	0.848
F	48	-0.121	0.965	0.071	0.834
G	75	0.367	0.849	0.069	0.767
H	30	-0.946	0.898	-0.735	1.246
J	80	0.019	0.969	0.007	0.941
K	29	0.130	0.666	0.083	0.783
L	24	0.348	0.770	0.647	0.752
M	6	0.839	0.867	0.694	0.584
N	29	0.163	0.577	0.087	0.489
P	29	0.606	0.505	0.293	0.881
R	9	0.411	0.299	0.327	0.587

TABLE E.3: Summary statistics for dimension 3, "Retrospecting vs. Notifying," scores by genre

		Brown Corpus		Frown Corpus	
Genre	n	Mean	SD	Mean	SD
A	44	0.147	0.715	0.446	0.730
B	27	-0.121	0.595	-0.278	0.613
C	17	-0.899	0.553	-0.740	0.678
D	17	-0.462	0.540	-0.672	0.644
E	36	-0.970	0.762	-0.614	0.925
F	48	-0.115	1.041	-0.474	0.646
G	75	0.012	0.945	0.121	1.012
H	30	-0.418	0.456	-0.368	0.764
J	80	-0.550	0.766	-0.596	0.827
K	29	1.055	0.772	0.528	1.000
L	24	1.176	0.921	1.090	0.826
M	6	0.889	0.633	1.146	0.741
N	29	0.852	0.653	0.661	0.492
P	29	0.849	0.717	1.055	1.211
R	9	0.278	1.007	0.917	1.085

TABLE E.4: Summary statistics for dimension 4, "Instructing," scores by genre

		Brown Corpus		Frown Corpus	
Genre	n	Mean	SD	Mean	SD
A	44	-0.101	0.721	-0.240	0.737
B	27	-0.460	0.653	-0.793	0.811
C	17	0.903	0.500	0.819	0.574
D	17	0.108	0.644	0.211	0.927
E	36	-0.931	1.759	-0.400	1.522
F	48	0.062	0.987	0.049	0.968
G	75	0.343	0.718	0.479	0.608
H	30	-0.709	1.547	-0.615	1.311
J	80	0.281	0.900	0.330	0.793
K	29	0.088	0.626	0.073	0.804
L	24	-0.397	0.481	-0.080	0.516
M	6	-0.096	0.291	-0.234	0.305
N	29	0.242	0.465	0.035	0.603
P	29	-0.010	0.603	-0.438	0.855
R	9	0.465	0.717	-0.503	1.317

TABLE E.5: Summary stats for dimension 5, "Referencing Positive Relationships," scores by genre

		Brown Corpus		Frown Corpus	
Genre	n	Mean	SD	Mean	SD
A	44	0.355	0.585	0.239	0.595
B	27	0.055	0.689	0.232	0.495
C	17	-0.730	0.743	-0.358	0.658
D	17	-0.073	1.121	0.197	0.939
E	36	-0.297	1.102	-0.375	0.940
F	48	0.028	0.835	0.088	0.828
G	75	-0.081	0.854	-0.246	0.870
H	30	0.316	0.677	0.722	0.660
J	80	0.632	0.736	0.728	0.705
K	29	-0.032	1.086	-0.786	1.072
L	24	-0.062	1.302	-0.002	0.771
M	6	-0.680	0.555	-0.284	1.086
N	29	-0.178	1.200	-0.233	0.568
P	29	-0.924	1.269	-0.985	1.219
R	9	-1.075	0.922	-0.711	1.284

Bibliography

Aitchison, J. (1994). *Words in the Mind* (2nd ed.). Oxford, UK: Blackwell Publishers.

Austin, J. L. (1975). *How to do things with words* (2nd ed.). Cambridge, Mass: Harvard UP.

Bakhtin, M. M. (1986). The problem of speech genres. In C. Emerson & M. Holquist (Eds.), *Speech Genres & Other Late Essays* (pp. 60–102). Austin: University of Texas Press.

Bazerman, C. (1994a). *Constructing Experience*. Carbondale: Southern Illinois UP.

Bazerman, C. (1994b). Systems of genres and the enactment of social intentions. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 79–101). London: Taylor and Francis.

Bazerman, C. & Russell, D. R. (Eds.). (2003). *Writing Selves Writing Societies: Research from Activity Perspectives*. Fort Collins, Colorado: The WAC Clearinghouse. Available: http://wac.colostate.edu/books/selves_societies/ [23 Mar 03].

Becker, A. L. (1991). A short essay on languaging. In F. Steier (Ed.), *Research and Reflexivity* (pp. 226–234). Thousand Oaks, Calif.: Sage Publications.

Becker, B. G. (1997). Using mineset for knowledge discovery. *IEEE Computer Graphics and Applications*, *17*(4), 75–78.

Benninghoff, S., Murphy, M., & Hart-Davidson, W. (2002). Should every student have an address on the digital street?: Electronic portfolios and outcomes-based assessment. Presentation at the 2002 CCCC.

Bereiter, C. & Scardamalia, M. (1987). Two models of composing processes. In *The Psychology of Written Communication*. Mahwah, NJ: Lawrence Erlbaum.

Berkenkotter, C. & Huckin, T. N. (1995). *Genre Knowledge in Disciplinary Communication*. Hillsdale, NJ: Lawrence Erlbaum.

Berlin, J. A. (1987). *Rhetoric and Reality: Writing Instruction in American Colleges, 1900–1985*. Carbondale, IL: Southern Illinois UP.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge, UK: Cambridge UP.

Biber, D. (1989). A typology of English texts. *Linguistics*, *27*(1), 3–43.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge Approaches to Linguistics. Cambridge, UK: Cambridge UP.

Birkerts, S. (1994). *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. Boston, Mass.: Faber and Faber.

Bollen, K. A. (2001). Indicator: Methodology. In N. Smelser & P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 7282–7287). Oxford, UK: Elsevier Science.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–634.

Booth, W. C. (1961). *Rhetoric of Fiction*. Chicago: Chicago UP.

Bracewell, R. J. (1999). Objects of study in situated literacy: The role of representations in moving from data to explanation. *Written Communication*, *16*(1), 76–92.

Bracewell, R. J. & Breuleux, A. (1994). Substance and romance in analyzing think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking About Writing: Reflections on Research Methodology* (pp. 55–88). Thousand Oaks, Calif.: Sage Publications.

Brush, A. J. B., Bargeron, D., Gupta, A., & Cadiz, J. J. (2001). Robust annotation positioning in digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 285–292). ACM Press.

Bryant, F. B. & Yarnold, P. R. (1995). Principal-components analysis and exloratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* (pp. 99–136). Washington, D.C.: American Psychological Assoc.

Bryson, B. (2001). *A Walk in the Woods: Rediscovering America on the Appalachian Trail*. New York: Broadway Books.

Bybee, J. & Hopper, P. (2001). Introduction. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 1–24). Philadelphia, PA: John Benjamins Publishing.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. San Francisco, Calif.: Morgan Kaufmann Publishers.

Carver, R. (1999). *Doing Data Analysis with MiniTab 12*. Pacific Grove, CA: Duxbury Press.

Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: U of Chicago Press.

Charney, D. H. & Carlson, R. A. (1995). Learning to write in a genre: What student writers take from model texts. *Research in the Teaching of English*, *29*, 88–125.

Chenoweth, N. A. (1997). *Reading for Revision : A Comparison of Teachers' and Students' Judgments*. PhD thesis, Carnegie Mellon University.

Chun, M. M. & Wolfe, J. M. (2001). Visual attention. In E. B. Goldstein (Ed.), *Blackwell Handbook of Perception* (pp. 272–299). Malden, Mass.: Blackwell Publishers.

Cicero (1954). *Ad Herennium: On the Theory of Public Speaking*. Loeb Classical Library. Cambridge, MA: Harvard UP.

Coe, R. M. (1990). Defining rhetoric—and us. *Journal of Advanced Composition*, *10*(1), 39–52.

Coe, R. M. (1994a). 'An arousing and fulfilment of desires': The rhetoric of genre in the process era—and beyond. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 181–190). London: Taylor and Francis.

Coe, R. M. (1994b). Teaching genre as process. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 157–169). London: Taylor and Francis.

Coe, R. M., Lingard, L., & Teslenko, T. (Eds.). (2002). *Rhetoric and Ideology of Genre: Strategies for Stability and Change*. Cresskill, NJ: Hampton Press.

Collins, J. (2002). Complex patterns and created experience: Building on a visualization technology for the writing classroom. Presentation at the Computers and Writing 2002 Conference.

Collins, J., Kaufer, D. S., Vlachos, P. K., Butler, B. S., & Ishizaki, S. (In Review). Detecting collaborations in text: Characterizing the authors' rhetorical language choices in the Federalist papers. *Computers in the Humanities*.

Conrad, S. (1999). The importance of corpus-based research for language teachers. *System*, *27*(1), 1–18.

Conrad, S. & Biber, D. (2001). Multi-dimensional methodology and the dimensions of register variation in English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-Dimensional Studies* (pp. 13–42). Harlow, England: Pearson Education Ltd.

Cooper, C. R. (1999). What we know about genres, and how it can help us assign and evaluate writing. In C. R. Cooper & L. Odell (Eds.), *Evaluating Writing: The Role of Teachers' Knowledge About Text, Learning, and Culture*. Urbana, Illinois: National Council of Teachers of English.

de Haan, C. B., Chabré, G., Lapique, F., Regev, G., & Wegmann, A. (1999). Oxymoron, a non-distance knowledge sharing tool for social science students and researchers. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, (pp. 219–228). ACM Press.

Downs, D. (2002). Representing gun owners: Frame identification as social responsibility in news media discourse. *Written Communication*, *19*(1), 44–75.

Efron, B. (1979). The bootstrap. In *The Jackknife, the Bootstrap and Other Resampling Plans* (pp. 27–36). Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B., Rogosa, D., & Tibshirani, R. (2001). Resampling methods of estimation. In N. Smelser & P. Baltes (Eds.), *International Encyclopedia of Behavioral and Social Sciences*, volume 19 (pp. 13216–20). Elsevier Science.

Elbow, P. (1994). Introduction. In P. Elbow (Ed.), *Landmark Essays on Voice and Writing*, volume 4 (pp. xi–xlvii). Davis, CA: Hermagoras Press.

Enos, T. (1990). An eternal golden braid: Rhetor as audience, audience as rhetor. In G. Kirsch & D. H. Roen (Eds.), *A Sense of Audience in Written Communication* (pp. 99–114). Newbury Park: Sage Publications.

Ericsson, P. & McGee, T. (2001). Squiggley green lines and red ink: Examining the "innards" of the Microsoft writing coach. Presentation at the 2001 Computers in Writing conference.

Fahnestock, J. (1993). Genre and rhetorical craft. *Research in the Teaching of English*, *27*(3), 265–271.

Faigley, L., Cherry, R. D., Jolliffe, D. A., & Skinner, A. M. (1985). *Assessing Writers' Knowledge and Processes of Composing*. Norwood, New Jersey: Ablex.

Fauconnier, G. & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fellbaum, C. (1998). *Wordnet : An Electronic Lexical Database*. Language, Speech, and Communication. Cambridge, Mass: MIT Press.

Flower, L. (1988). The construction of purpose in writing and reading. *College English*, *50*(5), 528–550.

Flower, L. (1994). *The Construction of Negotiated Meaning: A Social Cognitive Theory of Writing*. Carbondale: Southern Illinois UP.

Foertsch, J. (1995). Where cognitive psychology applies: How theories about memory and transfer can influence composition pedagogy. *Written Communication*, *12*(3), 360–383.

Francis, W. N. & Kucera, H. (1979). Brown corpus manual. Internet. Available: http://www.hit.uib.no/icame/brown/bcm.html [12/1/2002].

Freadman, A. (2002). Uptake. In R. M. Coe, L. Lingard, & T. Teslenko (Eds.), *The Rhetoric and Ideology of Genre* (pp. 39–53). Cresskill, NJ: Hampton Press.

Freedman, A. & Medway, P. (1994). Locating genre studies: Antecedents and prospects. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 1–22). London: Taylor and Francis.

Gage, N. (1991). The obviousness of social and educational research results. *Educational Researcher*, *20*(1), 10–16.

Gardner, C., Kirk, G., Inman, J., Hewett, B., & Maid, B. (2001). Hatching an OWL: Planning, developing, and maintaining an online writing lab. Presentation at the 2001 Computers and Writing conference.

Geisler, C. (1994). *Academic Literacy and the Nature of Expertise : Reading, Writing, and Knowing in Academic Philosophy*. Hillsdale, N.J.: Lawrence Erlbaum Assoc.

Geisler, C. (2001). Textual objects: Accounting for the role of texts in the everyday life of complex organizations. *Written Communication*, *18*(3), 296–325.

Geisler, C., Bazerman, C., Doheny-Farina, S., Gurak, L., Haas, C., Johnson-Eilola, J., Kaufer, D., Lunsford, A., Miller, C., Winsor, D., & Yates, J. (2001). ITEXT: Future directions for research on the relationship between information technology and writing. *Journal of Business and Technical Communication*, *15*(3), 269–308.

Gershon, N. & Eick, S. G. (1997). Information visualization. *IEEE Computer Graphics and Applications*, *17*(4), 29–31.

Haas, C. (1996). *Writing Technology: Studies on the Materiality of Literacy*. Mahwah, NJ: Lawrence Erlbaum.

Haas, C. & Flower, L. (1988). Rhetorical reading strategies and the construction of meaning. *College Composition and Communication*, *39*(2), 167–183.

Haller, C. R. (2000). Rhetorical invention in design: Constructing a system and spec. *Written Communication*, *17*(3), 353–389.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd ed.). London: Edward Arnold.

Halloran, S. M. (1990). From rhetoric to composition: The teaching of writing in America to 1900. In J. J. Murphy (Ed.), *Short History of Writing Instruction* (pp. 151–182). Davis, Calif.: Hermagoras Press.

Hanson, J. (1997). Sociality in the rhetorics of Kenneth Burke and Chaim Perelman: Toward a convergence of their theories. Paper presented at the Nat'l Communication Assoc. Conference. Available http://www.sla.purdue.edu/people/engl/dblakesley/burke/hanson.html [5 Oct 2002].

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17*(3), 307–352.

Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics, 8*(1), 9–20.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum.

Hayes, J. R. & Flower, L. S. (1994). Identifying the organization of writing processes. In R. E. Young & Y. Liu (Eds.), *Landmark Essays on Rhetorical Invention in Writing* (pp. 153–179). Davis, CA: Hermagoras Press.

Herrington, A. & Moran, C. (2001). What happens when machines read our students' writing? *College English, 63*(4), 480–499.

Hillocks, G. (1986). *Research on Written Composition: New Directions for Teaching*. Urbana, Illinois: National Council of Teachers of English.

Hillocks, G. (1995). *Teaching Writing as a Reflective Practice*. New York: Teachers College Press.

Hofland, K., Lindebjerg, A., & Thunestvedt, J. (1999). International computer archive of modern and medieval English (ICAME) collection of English language corpora. 2nd Edition (June 1999). Available: http://helmer.hit.uib.no/icame/cd/ [1/13/03].

Honeycutt, L. (2001). Comparing e-mail and synchronous conferencing in online peer response. *Written Communication, 18*(1), 26–60.

Hosmer, David W., J. & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.

Hundt, M., Sand, A., & Siemud, R. (1999). Manual of information to accompany the Freiburg–Brown corpus of American English (Frown). Info available: http://www.hit.uib.no/icame/cd [12/1/2002].

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, *60*(2), 237–263.

Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication*, *47*(4), 549–566.

Hutchins, E. (1996). Learning to navigate. In S. Chaiklin & J. Lave (Eds.), *Understanding Practice: Perspectives on Activity and Context* (pp. 35–63). Cambridge, UK: Cambridge UP.

Hyland, K. (2001). Bringing in the reader: Addressee features in academic articles. *Written Communication*, *18*(4), 549–574.

Hylton, J. (2000). Complete works of William Shakespeare. Internet. http://the-tech.mit.edu/Shakespeare/works.html.

Johnson, R. R. (1998). *User-Centered Technology: A Rhetorical Theory for Computers and Other Mundane Artifacts*. Studies in Scientific and Technical Communication. Albany, NY: SUNY Press.

Johnstone, B. (2002). *Discourse Analysis*. Introducing Linguistics. Malden, Mass.: Blackwell Publishers.

Kachigan, S. K. (1991). *Multivariate Statistical Analysis* (2nd ed.). FDR Station, NY: Radius Press.

Kaufer, D. S. (1981). Understanding ironic communication. *Journal of Pragmatics*, *5*(4), 495–510.

Kaufer, D. S. & Butler, B. S. (1996). *Rhetoric and the Arts of Design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kaufer, D. S. & Butler, B. S. (2000). *Principles of Writing as Representational Composition: Designing Interactive Worlds with Words*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kaufer, D. S., Hayes, J. R., Ishizaki, S., & Ishizaki, K. (2000). Using visualization technology to teach post-secondary writing. Final proposal to FIPSE, Carnegie Mellon University, Pittsburgh, Penn.

Kaufer, D. S. & Ishizaki, S. (2000). Docuscope update. Talk delivered for Human-Computer Interaction Institute Lecture Series, [8 Nov 00].

Kaufer, D. S., Ishizaki, S., Butler, B. S., & Collins, J. (In Press). *The Power of Words: Language and the Priming of Audience*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kaufer, D. S., Ishizaki, S., Ishizaki, K., Butler, B. S., Collins, J., Vlachos, P. K., & Ritivoi, M. (2002). Docuscope. Internet. Available: http://betterwriting.net/docuscope [1/12/03].

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1–12.

Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to Design & Analysis* (2nd ed.). New York: W. H. Freeman and Company.

Kiley, F. & Shuttleworth, J. M. (Eds.). (1971). *Satire: From Aesop to Buchwald*. New York: Odyssey/Bobbs-Merrill.

Kintsch, W. (1989). Learning from text. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction* (pp. 25–46). Mahwah, NJ: Lawrence Erlbaum.

Koedinger, K. R. & Anderson, J. R. (1993). Reifying implicit planning in geometry: Guidelines for model-based intelligent tutoring system design. In S. Lajoie & S. Derry (Eds.), *Computers as Cognitive Tools*. Hillsdale, NJ: Erlbaum.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, *8*, 353–388.

Lunsford, A. (2002). Technologies and college writers: The Stanford study of writing. Presentation at the 2002 CCCC.

Marcu, D. (2000). The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, *26*(3), 395–448.

May, J. M. & Wisse, J. (2001). Introduction. In *Cicero: On the Ideal Orator (De Oratore)* (pp. 3–56). New York: Oxford UP.

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech, 70*, 151–167.

Miller, C. R. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 67–78). London: Taylor and Francis.

Morris, J. H., Neuwirth, C. M., Regli, S. H., Chandhok, R., & Wenger, G. C. (1999). Interface issues in computer support for asynchronous communication. *ACM Computing Surveys (CSUR), 31*(2es), 11.

Neuman, W. L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches* (4th ed.). Boston, MA: Allyn and Bacon.

Neuwirth, C. M., Chandhok, R., Kaufer, D. S., Morris, J. H., Erion, P., & Miller, D. (2001). Computer support for distributed collaborative writing: A coordination science perspective. In T. W. Olson & J. B. Smith (Eds.), *Coordination Theory and Collaboration Technology* (pp. 535–558). Mahwah, NJ: Lawrence Erlbaum.

Neuwirth, C. M., Morris, J. H., Regli, S. H., Chandhok, R., & Wenger, G. C. (1998). Envisioning communication: task-tailorable representations of communication in asynchronous work. In *Proceedings of the ACM 1998 conference on Computer supported cooperative work*, (pp. 265–274). ACM Press.

Nickerson, C. (2000). *Playing the Corporate Language Game: An Investigation of the Genres and Discourse Strategies in English Used by Dutch Writers Working in Multinational Corporations*. Amsterdam: Rodopi.

Noden, H. R. (1999). *Image Grammar: Using Grammatical Structures to Teach Writing*. Portsmouth, NH: Boynton/Cook.

Norman, D. A. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Cambridge, Mass: Perseus Books.

Nystrand, M. (1986). *The Structure of Written Communication: Studies in Reciprocity between Writers and Readers*. Orlando, Florida: Academic Press.

Oakley, T. V. (1999). The human rhetorical potential. *Written Communication, 16*(1), 93–128.

Olson, D. R. (1994). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge, UK: Cambridge UP.

Ong, W. (1975). The writer's audience is always a fiction. *PMLA*, *90*, 9–21.

Palmquist, M., Carley, K. M., & Dale, T. A. (1997). Applications of computer-aided text analysis: Analyzing literary and nonliterary texts. In C. W. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts* (pp. 171–189). Mahwah, NJ: Lawrence Erlbaum.

Paltridge, B. (1997). Approaches to genre. In *Genre, Frames, and Writing in Research Settings* (pp. 5–46). Philadelphia, PA: John Benjamins Publishing.

Partington, A. (1998). *Patterns and Meanings : Using Corpora for English Language Research and Teaching*. Studies in Corpus Linguistics, V. 2. Amsterdam ; Philadelphia: J. Benjamins Pub.

Perelman, C. & Olbrechts-Tyteca, L. (1969). *The New Rhetoric: A Treatise on Argumentation*. Notre Dame, [Ind.]: University of Notre Dame Press.

Perfetti, C. A. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, *25*(2&3), 363–377.

Rauber, A. & Muller-Kogler (2001). Integrating automatic genre analysis into digital libraries. In *First ACM / IEEE-CS joint conference on digital libraries*, (pp. 1–10)., Roanoke, Virginia. ACM Press.

Richards, I. A. (1991a). Context theory of meaning and types of context. In A. E. Berthoff (Ed.), *Richards on Rhetoric* (pp. 111–117). New York: Oxford UP.

Richards, I. A. (1991b). Meanings as instruments. In A. E. Berthoff (Ed.), *Richards on Rhetoric* (pp. 190–199). New York: Oxford UP.

Roberts, C. W. (1997). *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Mahwah, NJ: Lawrence Erlbaum.

Roussinov, D., Crowston, K., Nilan, M., Kwasnik, B., Cai, J., & Liu, X. (2001). Genre based navigation on the web. Presentation at the IEEE International Conference on System Science.

Russell, D. R. (1995). Activity theory and its implications for writing instruction. In J. Petraglia (Ed.), *Reconceiving Writing, Rethinking Writing Instruction* (pp. 51–77). Mahwah, NJ: Lawrence Erlbaum.

Ryan, B. F., Ryan, T. A., & Joiner, B. L. (2000). Minitab. Version 13.32, Available: http://www.minitab.com [02/15/03].

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Schriver, K. A. (1997). *Dynamics in Document Design*. New York: John Wiley & Sons.

Scolari (2000). Diction 5.0: The text analysis program. Internet. Available: http://www.scolari.co.uk/diction/diction.htm [6/26/01].

Siegel, R., Woolsey, P., Correia, J., & Kalkwarf, S. (2002). Bbedit. Version 7.0.1, Available: http://www.barebones.com [01/10/03].

Sloane, T. O. (1989). Reinventing *Inventio*. *College English*, *51*(5), 461–473.

Spoerri, A. (1993). Infocrystal: A visual tool for information retrieval. In *Visualization '93*, (pp. 150–157)., San Jose, Calif.

Steuck, K. & Rowley, K. (2000). *Cognitive Tutor for Writers: Teacher's Guide*. San Antonio, Texas: TutorTek.com.

Steuck, K., Rowley, K., & Kretschmer, M. (1999). Partnering to implement computer-based tutoring systems in secondary schools. *Journal of Interactive Instruction Development*, *12*(1).

Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Stone, P. J. (2000). Comparison of General Inquirer with other text-analysis procedures. Internet. Available: http://www.wjh.harvard.edu/ inquirer/ [4/4/01].

Suthers, D., Lesgold, A., Katz, S., Weiner, A., Toth, E., Harrigal, K., Jones, D., Toth, J., Connelly, J., & DeLeon, Y. (1997). An interface for collaborative and coached approaches to learning critical inquiry. In *Proceedings of the 2nd International Conference on Intelligent User Interfaces*, (pp. 217–220). ACM Press.

Swales, J. M. (1990). *Genre Analysis: English in Academic Research Settings*. Cambridge, UK: Cambridge UP.

Swanson-Owens, D. & Newell, G. E. (1994). Using intervention protocols to study the effects of instructional scaffolding on writing and learning. In P. Smagorinsky (Ed.), *Speaking About Writing: Reflections on Research Methodology*, volume 8 of *Series in Written Communication* (pp. 141–162). London: Sage Publications.

Taylor, T. J. (1997). *Theorizing Language*. Amsterdam: Pergamon.

Tomlinson, B. (1990). Ong may be wrong: Negotiating with nonfictional readers. In G. Kirsch & D. H. Roen (Eds.), *A Sense of Audience in Written Communication* (pp. 85–98). Newbury Park: Sage Publications.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*, 99–114.

Van Dijk, T. A. (1986). News schemata. In C. R. Cooper & S. Greenbaum (Eds.), *Studying Writing: Linguistic Approaches* (pp. 155–185). Beverly Hills, Calif.: Sage Publications.

Vande Kopple, W. J. (1998). Relative clauses in spectroscopic articles in the Physical Review, beginnings and 1980: Some changes in patterns of modification and a connection to a possible shift in style. *Written Communication*, *15*(2), 170–202.

Vande Kopple, W. J. (2002). From the dynamic style to the synoptic style in spectroscopic articles in the Physical Review. *Written Communication*, *19*(2), 227–264.

Verheij, B. (1999). Automated argument assistance for lawyers. In *Seventh International Conference on Artificial Intelligence and Law*, (pp. 43–52)., Oslo, Norway. ACM Press.

Vlachos, P. (2001). Statlib datasets archive. Available: http://lib.stat.cmu.edu/datasets [10/21/02].

Wallace, D. L. (1991). *From Intention to Text: Developing, Implementing, and Judging Intentions in Writing*. Dissertation, Carnegie Mellon University, Pittsburgh, Penn.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches Tot He Communication of Knowledge*. Los Altos, CA: Morgan Kaufmann Publishers.

Wilks, S. S. (1932). Certain Generalizations in the Analysis of Variance. *Biometrika*, *24*, 471–494.

Winsor, D. (2000). Ordering work: Blue-collar literacy and the political nature of genre. *Written Communication*, *17*(2), 155–184.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., & Schur, A. (1995). Visualizing the non-visual: Spatiâl analysis and interaction with information from text documents. In *Proceedings of the Information Visualization Symposium 95*, (pp. 51–58)., Zurich. IEEE Computer Society.

Wittgenstein, L. (1958). *The Blue and Brown Books* (2nd ed.). New York: Harper Torchbooks.

Wolfe, J. L. & Neuwirth, C. M. (2001). From the margins to the center: The future of annotation. *Journal of Business and Technical Communication*, *15*(3), 337–371.

Wooden, J. A., DeVore, J., Graef, C., & Westman, L. (1998). Truth in advertising: Vintage cigarette and tobacco ads. Internet. Available: http://www.chickenhead.com/truth/index.html [2/10/03].

Wright, W. (1997). Business visuzlization applications. *IEEE Computer Graphics and Applications*, *17*(4), 66–70.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, *50*(3), 483–503.