

Thesis
1997
S165t

COPYRIGHT

By

Tânia Mara Gastão Saliés

December, 1997

TEXTS AS IMAGE-SCHEMAS: A CROSSLINGUISTIC
ANALYSIS OF DISCOURSE IN THE LIGHT OF
COGNITIVE LINGUISTICS

By

TÂNIA MARA GASTÃO SALIÉS

Bachelor of Arts
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, Brasil
1980

Master of Arts in English
Oklahoma State University
Stillwater, Oklahoma
1994

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirement for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1997

COPYRIGHT

BY

TaniaMara Gastao Sálies

December 1997

TEXTS AS IMAGE-SCHEMAS: A CROSSLINGUISTIC
ANALYSIS OF DISCOURSE IN THE LIGHT OF
COGNITIVE LINGUISTICS

Dissertation Approved:

Carl Lynn Mader

Dissertation Adviser

Cheryl M. Scott

Gene B. Halleck

Susan Gagnon

Dani Shevry

Wayne B. Powell

Dean of the Graduate College

*“Natural language possesses some inherent logic of its own,
a logic whose nature can only be revealed
by continued investigations.” (Quine)*

ACKNOWLEDGMENTS

This JOURNEY began years ago when I took a contrastive analysis class as a core undergraduate course at Pontifícia Universidade Católica do Rio de Janeiro. My interest for languages and the mapping of experience across different codes became then a constant topic in my research interests. Be it because of my work as a translator and interpreter, be it because of my work as a researcher and EFL teacher more recently. Seventeen years, 60 credits, numerous research projects, teaching hours, and knowledge crisis later, I get to my final destination—the completion of this dissertation and, ultimately, this degree. Because the journey has been long, friends, professors, family members contributed, in different ways, to its successful completion. To all of you I express my gratitude.

I owe special thanks to my adviser, Dr. Carol Moder, who continuously inspired, encouraged, and supported me along this JOURNEY. She introduced me to the world of cognitive linguistics and discourse analysis as well as guided me toward the completion of this study. Being such an exemplar role model, I feel fortunate to have had the opportunity to work with her. Also, I should like to thank Professor Paulo Brito who taught me that contrastive analysis course so many years ago. The text that follows develops many of the seeds he has planted. I should also thank the other members of my committee who took their time to read and comment on my work—Dr. Sheryl Scott, Dr. Gene Halleck, Dr. Susan Garzon, and Dr. Ravi Sheorey. My gratitude goes to Dr.

Richard Batteiger too. Dr. Batteiger kindly invited me to work for the Mid-level Writing Assessment project that financed much of my work at this University for almost two years, making thus this journey financially possible. Much of what I have learned about writing and the intricacies of composing, I owe to him. Our long discussions about students' writing and texts fed many of the ideas that I develop here. Finally, I would like to thank my husband Jacques for being such an exceptional friend and source of encouragement. His unconditional and unselfish support has been key to the completion of this degree. At times of frustration and discouragement, I only kept going because he urged me to. To him, I dedicate this dissertation and the Ph.D. is a JOURNEY metaphor.

TABLE OF CONTENTS

Chapter	Page
I. PROLOGUE	
Theoretical framework	5
Research goals	9
Organization of the dissertation	9
II. EXISTING ACCOUNTS OF WRITTEN DISCOURSE	
Overview	12
Texts as functional structures	13
Texts as knowledge structures	27
Texts as situated interaction	30
Texts as pragmatic inferences	41
Texts as linguistic, social, and cultural phenomena	46
Text as translatable relations in families of resemblances	62
Summary	66
III. THE COGNITIVE APPROACH	
Overview	69
Principles of categorization	70
Understanding image-schemas	74
Linguistic devices, discourse, and cognitive linguistics	77
Aspects of Cognitive Grammar	90
Linguistic relativism and the cognitive approach	101
Summary	107
IV. DISCOURSE PROCESSING	
Overview	109
As related to memory	109
Models of discourse processing	118
As related to linguistic devices	126
As related to other cognitive factors	137
As related to perception	140

Some preliminary conclusions	146
The COMMUNICATIVE TEXT image-schema	146
V. FACTS ABOUT BRAZILIAN PORTUGUESE	
Overview	153
Word order	153
Grammatical markings	156
Head-modifier relations	158
Summary	161
VI. THIS STUDY	
Overview	162
Rationale	162
Facts about the study	164
The "communicative text" category	164
Method	167
Data	167
Procedures	176
Data Analysis	177
Coding of the linguistic events	178
VII. TEXTS AS IMAGE-SCHEMAS: THE COMMUNICATIVE TEXT CATEGORY	
Overview	203
In English	
Quantitative Results	203
Sentence Length	203
Length of Attention Unit	206
Type-token ratio	208
Content Words	209
Nominalizations	211
Sequences of juxtaposed PPs	212
Adverbial expressions of time and place	214
Independent Constructions	214
Summary of quantitative results	215

Explaining the COMMUNICATIVE English TEXT	217
How the lexicon cues comprehension	218
How syntax cues comprehension	224
How English interacts with reality	234
Summary of qualitative analysis	236
In Brazilian Portuguese	
Quantitative Results	237
Sentence Length	237
Length of Attention Unit	240
Type-token ratio	243
Content Words	244
Nominalizations	246
Sequences of juxtaposed PPs	247
Adverbial expressions of time and place	250
Independent Constructions	250
Summary of quantitative results	251
Explaining the COMMUNICATIVE Brazilian Portuguese TEXT	253
How the lexicon cues comprehension	253
How syntax cues comprehension	263
How BP interacts with reality	278
Summary of qualitative analysis	280
Variation within the COMMUNICATIVE TEXT image-schema	282
Comparing Brazilian Portuguese to English	288
Syntactic Organization	290
Lexical Organization	294
Interaction with reality	295
Qualitative analysis	295
Summary	299
Summary of findings	331

Discussion	
Introduction to the discussion	332
Results in the light of the literature review	
BP discourse features in the light of existing studies	338
Attention units as measures of production and comprehension	341
Overriding differences between speech and writing	343
Research on genre	344
Linguistic Relativity	345
Results as linguistic, social, and cultural phenomena	347
Translation studies	348
General comments	351
 VIII. CONCLUSION	
Theoretical conclusions	352
Methodological conclusions	354
Implications	356
Summary of contributions	358
Recommendations for future studies	359
Conclusion	361
 REFERENCES	362
 APPENDICES	400
APPENDIX A— GLOSSARY	401
APPENDIX B— SAMPLE OF TEXTS IN THE CORPUS	408
APPENDIX C— A BP TEXT WITH GLOSS	414
APPENDIX D— BP TEXTS AND THEIR ENGLISH TRANSLATION VARIANTS	418

LIST OF TABLES

Table	Page
1. Data Profile	168
2. Linguistic Events	180
3. Sentence length in English	204
4. Most used sentence length in English	205
5. Length of attention units in English	207
6. Most used attention units in English according to length	207
7. Distribution of the type-token ratio across English texts	209
8. Number of content words per 1,000 words in English texts	210
9. Number of nominalizations per 1,000 words in English texts	211
10. Sequences of PPs per 1,000 words in English discourse	213
11. Locative/time adverbials per 1,000 words in English texts	214
12. Independent clauses per 1,000 words in English texts	215
13. Events associated with English texts	226
14. Sentence length in Brazilian Portuguese	237
15. Most used sentence length in Brazilian Portuguese	239
16. Length of attention units in Brazilian Portuguese	241
17. Most used attention units in BP according to length	243

Table		Page
19.	Distribution of the type-token ratio across BP texts	244
20.	Number of content words per 1,000 words in BP	245
21.	Number of nominalizations per 1,000 words in BP	246
22.	Sequences of PPs per 1,000 words in BP	248
23.	Typicality of prepositional phrase sequences per 1,000 words in BP	249
24.	Locative/time adverbials per 1,000 words in BP	250
25.	Independent clauses per 1,000 words in BP	251
26.	Events associated with Portuguese texts	252
27.	Comparing BP to English institutional discourse	289
28.	Comparing the rank sums of attributes in BP and English discourses	296
29.	English translation as a metaphor of English and BP discourses	301
30.	Comparing the three research conditions	303

LIST OF FIGURES

Figure		Page
1.	Schematic representation of basic grammatical categories according to CG	92
2.	Word order is an [ENERGY CHAIN] metaphor	94
3.	Nominalization according to the notation of CG	97
4.	A socio-cognitive model of the writing process	117
5.	Texts as image-schemas: The COMMUNICATIVE TEXT category	150
6.	The PATH sub-component structure	151
7.	Romance languages: Word-order-dependency continuum	160
8.	Getting at an image-schema of English and BP institutional discourse	179
9.	Sentence length in English	206
10.	Relative frequency of attention units as a function of number of words in English	208
11.	Ratio of content words in institutional English discourse	210
12.	Relative frequency of nominalizations per 1,000 words in English discourse	212
13.	Conceptualizing syntagmatic relations in English	229
14.	Sentence length in Brazilian Portuguese	240
15.	Relative frequency of attention units according to number of words in BP	242
16.	Typicality of content words in institutional BP discourse	245

Figure		Page
17.	Occurrence of nominalizations in BP	247
18.	Conceptualizing syntagmatic relations in BP	275
19.	The English COMMUNICATIVE TEXT image-schema	285
20.	The Brazilian Portuguese COMMUNICATIVE TEXT image-schema	287
21.	Sentence and attention unit lengths in BP and English discourse	290
22.	Comparing syntactic organization in English and BP	293
23.	Lexical organization in English and BP institutional discourse	294
24.	Comparing the mean rank sums in three research conditions	306
25.	Comparing the organization of syntax in BP and English translation	310

ABBREVIATIONS

m	masculine
s	singular
f	feminine
pl	plural
p	person
art	article
ind	indicative
subj	subjunctive
pret	preterit
pres	present
fut	future
Adj	adjective
Adv	adverb
N	noun
V	verb
NP	noun-phrase
VP	verb-phrase
PP	prepositional phrase
Adv.P	adverbial phrase
Pass	passive
TR	trajector
LM	landmark
Inf	infinitive
Ger	gerund
Part	participle

CHAPTER I

PROLOGUE

In the last two decades, research in the field of cognitive linguistics has seen voluminous growth (Coleman and Kay, 1981; Lakoff, 1990; Taylor, 1989; Langacker, 1987; 1991; Casad, 1996; among others). The same holds true for studies in the field of discourse analysis (Halliday and Hasan, 1976; Biber, 1988, 1995; Grabe, 1990; Brown and Yule, 1983; Schiffrin, 1995), pragmatics (Davis, 1991; Blackmore, 1992; Sperber and Wilson, 1986; 1995; Grice, 1975; 1989), and discourse processing (Tabasso, Secco & van den Broek, 1984; Smith and Swinney, 1992; Baddeley, 1990; Kintsch, 1995). However, investigations that combine the premises of these four fields to analyze texts are lacking. Also, though crosslinguistic literature that addresses discourse is abundant on languages such as Italian, French, German, Spanish (see Bates and MacWhinney, 1989; Connor, 1996; Grabe and Kaplan, 1996 for an overview), little work has been done with Portuguese (Moragne e Silva, 1991), particularly, Brazilian Portuguese (Almeida, 1984; Dantas-Whitney and Grabe, 1989; Oliveira, 1997a; 1997b).

From a linguistic perspective, the value of a study that would combine these fields of inquiry seems clear to me. First, they would be able to account for how categories that do not exist objectively in the world such as “texts,” “head-modifiers,” “inflections” or “metaphors” are grounded in daily experience and in the human mind (Lakoff, 1990; Casad, 1996). Second, they would provide insights into how conceptual knowledge

interacts with language in use. Categories that are part of our concrete reality should be easier to construe and understand. Third, investigations of that kind seem to be a powerful tool to reveal what is cognitively the same among languages and that which is language specific. Casad (1996) and Hunt and Agnoli (1991), for example, re-address the so much discussed Sapir-Whorf within a cognitive paradigm. Finally, studies that would follow such an interdisciplinary framework give researchers the opportunity to blend notions largely discussed by discourse analysts such as the background and foreground distinction with those discussed by philosophers of language such as relevance, felicity conditions; with those discussed by researchers interested in the mind and its workings such as cost of processing, cue and ecological validity, salience, consciousness; with those long discussed by descriptive linguists such as word order, SV agreement, inflections, and other grammatical constructions.

In this study, I will take up the challenge of combining those notions to analyze institutional expository discourse published in English, Brazilian Portuguese (BP), and the English translation of the BP texts. The sub-genre “institutional” refers to texts published by business companies to promote their images and businesses. Though the study is basically a theoretical one, I also intend to show quantitatively that discourse in English and Portuguese has certain properties which, when combined, form a central, basic category that is inclusive and exhibits a prototype structure (Lakoff, 1990). Because it is inclusive, it is simple and typically used, being therefore easier to understand than the extensions motivated by it.

To form a basic discourse category, writers blend pragmatic, discursive, cognitive, and perceptual considerations guiding readers to retrieve the corresponding image-schema

they intended to construe during invention. An image-schema, according to Lakoff (1990), is part of an idealized conceptual model (ICM) that is independent of the words, morphemes, and rule orders of particular languages but that is intrinsically connected to motor-perceptual experiences and to the nature of thought. That is to say that an image-schema is part of a universal system of cognitive representation (Langacker, 1991) or a supralinguistic system of knowledge (Mc Cormack, 1977; Potter, So, Von Eckardt, & Feldman, 1984).

The rationale for such an approach rests on the need for a theory of discourse analysis that links competence with performance, language production with cognition, cognition and verbal behavior with the sociocultural context. Only such an interdisciplinary viewpoint may do justice to the overwhelming intricacies of language.

The link between language and cognition has been supported by existing experimental studies in a variety of languages (Bates and MacWhinney, 1989; Hoover, 1992; Shridhar, 1989). Those studies have demonstrated that language users rely on linguistic (word order, animacy, SV agreement) and cognitive (cue and ecological validity, prominence) cues to construe meaning. English speakers, for example, rely primarily on word order while speakers of morphologically rich languages such as Italian and Spanish rely on SV-clitic agreement, animacy, and word order. Also, this body of research indicates that that which is perceptually more salient tends to appear first in sentence production or event retelling across languages (Shridar, 1989; Flores d'Arcais, 1987). Another important result deriving from this body of research shows that English discourse tends to have shorter sentences than Italian and Spanish and to favor a late-closure

processing strategy (Hunt and Agnoli, 1991; Hoover, 1992). In Enkvist's (1987) words: "Languages look the way they do because they have to be processed..." (p.40)

In the light of that body of knowledge, I theorize that discourse is what I call COMMUNICATIVE TEXT image-schema. The COMMUNICATIVE TEXT image-schema is an abstract structure that embodies the commonalities of different schemas—SOURCE-PATH-GOAL-DESTINATION (Lakoff, 1990)—as well as of the different discourse variants that define those schemas in different and contrasting dimensions. Many attributes of the category will be true of many members, but there are some attributes that will be true of only some members. Those cases that are more specific, inclusive, and readily identified are more prototypical and serve as a cognitive reference point.

In addition, the attributes that cluster together to define the COMMUNICATIVE TEXT category are not inherent to institutional expository discourse. Rather, they *are interactional properties* (Lakoff, p.51), and relate to the way conceptualizers perceive, imagine, organize, and behave toward that combination of attributes (Rosch, 1978). Thus, the cluster acts as a whole, or gestalt, that psychologically is simpler than its parts.

Because of the descriptive similarity and historical relation between Portuguese, Spanish, and Italian and existing research studies, I hypothesize that a Portuguese COMMUNICATIVE TEXT will show attributes like lengthier sentences and more nominalizations and right-branching constructions than English. Those linguistic features mark underlying functional dimensions, as Biber (1988) says in the context of his research on linguistic variation across texts. Linguistic features do not randomly co-occur across texts. Their clustering is motivated and grounded in experience. The specific motivation

this dissertation will address is effort of processing. I would like to propose, for example, that those features co-occur in Brazilian Portuguese because of BP's rich morphology.

Verb, noun and adjective inflections function as anaphora in BP, allowing on-line processing and balancing the processing load brought to bear by lengthy sentences.

Theoretical Framework

This study stems from a theory of language that views linguistic phenomena as a product of various cognitive mechanisms (Gibbs, 1996; Langacker, 1987; Lakoff, 1990; and Taylor 1989), and discourse specifically as a process and not merely a set of structures (Enkvist, 1987; Chafe, 1982; 1987; 1994).

The first theoretical assumption guiding its design is that language is a non-autonomous, motivated system that interacts with knowledge structures, cognitive skills, and sensory experience to accomplish its fundamental function: to communicate. In linguistic analyses that follow such a line of thought, meaning derives from pragmatic and discourse related matters, reflecting categories that are present in the human mind (Lakoff, 1990) and that are responses to the socio-cultural environment.

Second, the study is based on a theory of discourse that defines written texts as interactive speech events (Hymes, 1972; Basso, 1974; Widdowson, 1979; Scribner and Cole, 1981; Tannen, 1982b; Nystrand, 1987; Carrell, 1987; Tadros, 1994) between an addresser (the writer) and an invisible addressee (the reader)—“a discourse-as-process view” (Brown and Yule, 1983, p.24). As such, it is situated and part of a socio-cultural context. Linguistic devices become then writers' instruments toward a major objective: “communication”. To convey an intended image-schema, writers must incorporate a multiplicity of linguistic (rhetorical structure; word order; vocabulary; morphological

trappings) and non-linguistic (culture, social and gestalt norms) cues into their speech events. If one fails, successful communication is jeopardized. Take for example, constructions that offset the prominence of a dominant idea. One such case is the presence of a main idea in a relative clause. Regardless of other situational and linguistic cues, such structures are prone to ambiguity (see Tomlin, 1985; Erschik-Shir and Lappin, 1979; Saliés, 1995d) and force different readers to fill in the gaps distinctively. By using that type of construction writers divert readers from what is relevant for comprehension.

Third, as I have mentioned, this dissertation adopts an interdisciplinary research strategy, incorporating findings from disciplines such as psycholinguistics, cognitive psychology, sociolinguistics, and discourse analysis. Though I do not conduct on-line studies in sentence processing or collect live data for description, I use evidence from that literature to refine our understanding of how comprehension relates to the use of linguistic devices in natural written speech event settings—published discourse. Most of all, I assume that natural written speech events reflect processes in writers' minds and anticipate those that may be in readers' minds.

The linguistic analysis will follow notions of cognitive grammar (Langacker, 1987; 1991). Among those notions I highlight:

- (1) The base-profile distinction
- (2) The degree of prominence scale
- (3) The setting-participant asymmetry
- (4) The ability to shift mentally from one domain to the other
- (5) The subjectivity-objectivity distinction

Fourth, the study calls on pragmatic theory as advanced by Sperber and Wilson (1986; 1995; Wilson and Sperber 1991), who built on Grice's communication cooperative principles and maxims of truthfulness, informativeness, relevance and brevity. According to Sperber and Wilson (1995) "every act of ostensive communication communicates a presumption of its own optimal relevance" (p. 260), a maxim known as *The Principle of Relevance*. That is, the interpretation of discourse involves pragmatic processes that are highly context dependent and that yield contextual effects that constrain readers' interpretations to the one intended by the addresser, at a minimal processing cost.

Last, in describing discourse in Portuguese and English I will call on concepts of semantic theory such as markedness and blocking. Markedness is a term used by linguists to describe binary contrasts or as Lakoff (1990) says, asymmetries established by the presence of additional morphologic or semantic material within a category. The unmarked member of a category is the usual, more frequent, and less complex form structurally and cognitively (the more basic member of the category). The marked member is the less common and frequently used, a more complex form. In English, singular number is the unmarked form within the number category because it is more frequent, simpler, and easier to process than plural, the counter-norm. Actually, unmarked members constitute the culturally accepted norm since language users perceive that which is common as positive and that which is infrequent as negative. Until not long ago, for instance, being a female lawyer carried a negative connotation because it was not the norm to find a female lawyer. This fact was true of the US, Brazil, and other locations across the globe. The markedness concept is especially useful in this study to help us understand why certain forms are easier to process than others, and why the COMMUNICATIVE TEXT

category carries a positive connotation. Actually it blocks cases that yield low cognitive effects.

Blocking is a term linguists use to explain lexical choices that view maximum communication efficiency. If I choose a word that is tightly restrictive in its semantic atom, I choose it because it carries that extra semantic information that is necessary for the addressee to understand the message fully in an optimum time. Such a word blocks others of more general meaning (e.g. animals, dogs, Dobermans). Both markedness and blocking are types of prototype effects; they structure the linguistic system in hierarchical levels of complexity, being relevant for the efficient completion of the speech act (see Hoffman, 1993).

A note about the theoretical framework and terminology is in order. Following Connor (1996), and Virtanen (1990), this study treats text analysis as synonymous with discourse analysis or discourse linguistics, a superordinate cover term proposed by Enkvist (1987). That terminology has been used with different connotations in the literature. The distinction among the terms has to do with the language channel (written or spoken) and the approach (looking or not at interactional and situational contexts) of a study. While analyses of written language that do not take into account communicative constraints have been called text linguistics, analyses of spoken language that examine the socio-cultural and psychological universes of language users have been known as discourse analysis. However, as I mentioned, written language is essentially a speech event (Hymes, 1972; Basso, 1974), and analyses of speech events cannot truly exist as such if they fail to consider the communicative constraints of the situation: "The text itself provides a context for its constituent parts" (Enkvist, 1987, p.26).

Research Goals

Specifically, the research goals that guide this study are the following:

1. To analyze the distribution of 8 linguistic devices across 20 institutional expository texts in Portuguese, English, and English translation. Namely, I analyze sentence and attention unit length, the type-token ratio, occurrence of content words, nominalizations, location and time adverbials, sequences of prepositional phrases, and independent clauses.
2. To account for how these linguistic categories relate to cost of processing using the tools of Cognitive Grammar and Relevance Theory.
3. To provide and compare a schematic characterization of the COMMUNICATIVE TEXT in BP and English.
4. To demonstrate how this schematic characterization of texts cover communicative needs.
5. To explain how this type of analysis contributes to a theory of discourse analysis and of translation in line with the cognitive approach.

Results of such analysis should be of particular interest to linguists in general, psycholinguists, cognitive psychologists, writing specialists, translators, and researchers interested in cross-linguistic studies.

Organization of this dissertation

This study is organized into eight chapters. To build on existing linguistic theory, I introduce the reader to the related literature on discourse analysis, the cognitive approach, and sentence and/or discourse processing (Chapter II, III, and IV respectively) and detail facts about the Portuguese language (Chapter V). Because I establish the theoretical

premises of the study and offer new theories in these four theoretical chapters, they constitute the backbone of this dissertation. In Chapter IV, for example, I propose and explain the COMMUNICATIVE TEXT category and in Chapter V, I briefly compare facts about Portuguese to facts about English and suggest a continuum for word-order dependency. Therefore, the organization of this dissertation departs from other studies in what has been typically associated with a model of literature review. In this study, the literature review not only sets the study into a broader context, but also offers new theoretical notions.

Moreover, the chapters that review the literature were organized in that order to flow from information that is generally most known to information that is generally least known to readers in the field. In a way, I tried to make the organization of chapters a metaphor of the given-new contract, saving for last (Chapters III and IV) information that constitutes the meat of the descriptive tools I use to analyze the corpus—the cognitive paradigm and discourse processing.

Chapter VI deals with the methodology and relevant literature, specifying the type and number of texts present in the study and how they were analyzed quantitatively and qualitatively. Chapter VII presents the results of the analysis, describing the COMMUNICATIVE TEXT category in English and Brazilian Portuguese, comparing the two, explaining discourse translation cognitively, and showing how the results pertain to the research goals. The conclusion (chapter VIII) integrates major findings with existing research, projects pedagogical and theoretical implications, and suggests future venues for research that would follow a similar framework. To support the methodological

discussion and allow further consideration of the analysis, the appendices offer a sample of specific texts from the corpus as well as a glossary of terms frequently used in the study.

To close this prologue, I would like to emphasize that this is not a conventional empiric study. The quantitative and qualitative analyses simply provide evidence for theoretical notions that I develop in chapters I, II, III, IV, and V. These chapters not only establish the theoretical basis of the study, but also offer new notions. Actually, they ground the discussion and the conclusion of this work.

CHAPTER II

EXISTING ACCOUNTS OF WRITTEN DISCOURSE

Overview

Researchers from a variety of fields and theoretical perspectives have contributed to our current understanding of how to analyze written discourse. Insights from translation studies (Toury, 1991; Genzler, 1993; Wilss, 1988; Snell-Hornby, 1988) and investigations that characterized speech and writing (Halliday, 1987;1989; Chafe 1982; Chafe and Danielewicz, 1987; Chafe, 1987; Biber, 1988), recent research in discourse analysis (Grabe, 1996; Chafe, 1994; Shrifin, 1995; Givón, 1995; Grabe 1996), contrastive rhetoric (Connor, 1996; Kaplan, 1988; Grabe and Kaplan, 1989; Grabe, 1987) and reading theory (Carrell, 1987, 1992) combine to show that written discourse is a speech event with a “complex multidimensional structure” (Kaplan, 1988; Biber, 1988), including schemas which are conceptual and pragmatic such as experiential gestalts, inferences, frames, topic, and patterns of genre development.

This chapter reviews that literature, discussing particularly information that relates directly to the scope of this dissertation: how linguistic attributes distributed across a text relate to perceptual-cognitive constraints. Instead of revealing the diversity and different approaches to the analysis of written discourse, I present features and conceptual principles that are critical to the understanding of the research on which I will report. For a comprehensive theoretical and practical review of the various discourse analysis

frameworks, I direct the reader to Schrifin (1995).

The review begins with studies that examined the distribution of linguistic properties in spoken or written discourse, limiting their analysis to the empirical observation of data and functional explanations. It continues with studies that explain the use of linguistic features vis-à-vis their import in the potential interaction between writers and readers. Among those studies is the work of researchers such as Chafe (1994), who has shown that successful study of discourse depends on introspective insights, not solely on the empirical observation of data. Third, the discussion approaches studies that addressed written discourse crosslinguistically bringing the thorny language-culture-thought discussion to light. Fourth, I address research on translation theory and its contribution to the understanding that though people share grammatical knowledge of a language, they contextualize it differently (Gumperz, 1982; Jackendoff, 1994). No analysis of written language can escape the fact that language is situated in culture and social life (Hymes, 1972), being but one of the components of the communication process and having, at the same time, communication as its ultimate objective (Shrifin, 1995). Finally, this section focuses on the general assumptions that writers and readers bring to each and every occasion of writing or the cooperative nature of participants in a written speech event (Grice, 1975; Sperber and Wilson, 1986; 1995; Blass 1990). In closing this chapter, I summarize the principles that emerge from the body of research here reviewed and that will guide my choice and analysis of data.

Texts as functional structures

Studies under this heading place linguistic categories into a framework of textual analysis, identifying recurrent patterns and regularities in discourse and examining how

such regularities serve the function of communication. Specially, the studies examine the position of selected linguistic units (morphologic, syntactic or meaning-based) in the sequential flow of information and ground them in a speech situation. That is, these studies see linguistic units as codes that addressers (writers) combine in a context (sequence of referring terms) to convey messages to addressees (readers). In addition, studies that match structure to function deal with “overt manifestations of language” (Chafe, 1994, p. 12) or with language that is publicly observable and natural (published texts). I review those studies according to the development of concepts and theories in discourse analysis, placing them in a rough chronological order.

Early studies had the sentence as the unit of investigation and focused on the theme-rheme asymmetry as proposed by the Prague school of linguistics (Halliday, 1967; Danes, 1974; Firbas, 1992). While the theme has been defined as information previously introduced in discourse, rheme became known as information that has just been newly introduced. The Prague School’s pioneer functional sentence perspective focused primarily on the linguistic factors characterizing the move from a theme to a rheme and on the degree of communicative dynamism involved in it. According to the School, the movement is made clear by linear modification of words in a sentence, semantic, and contextual aspects (identifiability in previous sequences), and, in the case of spoken language, prosodic prominence. For example, Prague linguists found that in unmarked sequences, the theme or information that can be identified in previous discourse, comes normally before the rheme. When the rheme or new information shows first, or early, it is marked somehow by the presence of a distinctive feature such as the use of explicit and indefinite expressions in clause-initial position in written language or high-pitch in spoken

language. Those moves in word order and prosody signal the development of speech acts toward the goal of the speech event: communication. In other words, the Prague School functionally explain word order.

Like the Prague School, Halliday (1967, 1970) has paired linguistic units to their functions (1978), posing that the three major functions of language—ideational, interpersonal, and textual—should be the backbone of any analysis that purports to investigate how sentences make up recognizable texts or unified messages (as opposed to strings of unconnected sentences).

For Halliday (1978), texts are sequences of *information units*, an obligatory new element (the message) plus an optional element that is retrievable from the previous text or context. The order in which those units appear in discourse depends primarily on the state of knowledge of the addressee. If writers choose to place an information unit first (the theme in Halliday's conception), it is because it connects back and links to previous discourse, serving as a point of departure for further development of language. In this sense, though a theme will not always be the topic or what the sentence is about, it will always reflect presupposed information or some kind of linkage between given and new elements (see Halliday 1985).

Studies began then to devote attention to those linguistic units that tie previous to following discourse—mechanisms of co-reference or cohesion (Halliday and Hasan, 1976; Hasan, 1978; Halliday and Hasan, 1989)—and that represent the textual function of language. Halliday and Hasan (1976) identified five major types of cohesive constructions: referential (pronouns and pronominal adverbs), lexical (repetition and collocation), substitution (pro-forms and zero-substitution or ellipsis), and conjunctive

cohesion. Those overt constructions help readers connect lexically and grammatically the theme to the rheme.

However, Halliday and Hasan's (1976) taxonomy has been criticized as a sole mechanism of textuality (Carrell, 1982; de Beaugrande and Dressler, 1981; Witte and Faigley, 1981). For example, critics reprove the absence of contextual elements such as purpose, time, place and genre of discourse as sources of "texture." Critics also claim that knowledge of the world is fundamental to "texture" construal and that readers may construe "texture" despite the absence of cohesive ties.

Regardless of the criticisms, Halliday and Hasan's (1976) taxonomy of cohesive constructions became a tool in investigations of discourse organization. Numerous studies in the field used it as the basis of their analysis (see Akinaso, 1982). The importance of their taxonomy, in my understanding, derives not only from the ease with which it can be applied to the analysis of texts in different languages but also from the inevitable fact that linguistic structures reveal many of the semantic intentions of the writer. In a recent study about coherence, Grabe (1996) reinforces that argument saying that "the surface structure of texts correlates strongly with the underlying textual coherence and...text coherence can be generated, in good part, from the text itself" (p.5).

Halliday's investigations (1967, 1970, 1979, 1987, 1989) have also revived the notion that spoken and written discourse have different properties, inspiring other studies that brought major contributions to this research (Chafe, 1982; Tannen, 1982^a; 1982^b; Chafe and Danielewicz, 1987; Biber, 1988). Because this dissertation is concerned with written discourse, my discussion of that body of knowledge places greater emphasis on findings related to writing, and given the similarities between the two modalities, it applies

findings related to spoken language to written discourse (the discussion that follows will treat those similarities).

Halliday's investigations (1987, 1989) have demonstrated that both language modalities are complex in their own way. While written language shows lexical complexity (twice as many lexical items as grammatical words per clause), spoken language shows grammatical complexity (more clauses or complex sentence structures). His studies have also highlighted the overwhelming frequency of nominals—nouns and their pre-modifying and post-modifying elements—in written English. In his view, that characteristic derives from the structure and function of the nominal group in English (pre-modifiers + head + post-modifiers) and the information structure of the clause (a theme with a topical element plus a rheme). Because the topical element within a theme is a nominal (the “phenomenon being referred to,” the “meat of the message”), it allows further embedding or post-modification by means of prepositional phrases or other nominal groups. Actually, it works thematically to establish footing and organize information.

Though Halliday (1987, 1989) recognizes that overuse of nominals may cause loss of ideational information and ambiguity, he stresses the gain in textual information they bring. According to his analysis, written English is a language with high lexical density. That is, English has a strong tendency to carry many lexical items per clause and to encode this lexical content in a nominal form: in head nouns, nouns, and adjectives in the nominal group, and nominalised clauses (1989, chapter five).

Chafe (1982), Tannen (1982b), Chafe and Danielewicz (1987), and Biber (1988) further characterized spoken and written language showing the distribution of categories

across existing corpora. General results from these studies indicate that written discourse is structurally more elaborated and explicit than speech, with longer sentences and greater use of subordination. Furthermore, those studies indicate that written language is more de-contextualized and less dependent on shared situation than spoken language. "Whereas written language fosters the kind of detachment evidenced in the use of passives and nominalizations, spoken language shows a variety of manifestations of the involvement which a speaker has with his or her audience" (Chafe, 1982, p.48).

Chafe (1982) reinforces Halliday's (1987, 1989) previous findings about nominalizations in a study in which he compared the frequency of occurrence of linguistic features in dinner table conversations and in academic papers. Nominalizations and attributive adjectives showed as the most prevalent features of written discourse. There were eleven and a half times as many occurrences of nominalizations and four times as many occurrences of attributive adjectives (occurrences per thousand words) in the written data when compared to the spoken. Contrary to Halliday (1987, 1989), Chafe sees those differences as a result of the extra time writers have to package information and not as a consequence of the structure of English and its textual function.

However, a closer look at that body of knowledge indicates that important variations occur across genres and that spoken and written discourse may be similar in different ways, mainly with respect to context and interaction. Tannen (1982b) observes that features such as high context dependency are as much true of conversations as they are of written genres such as personal letters or office memos. If we look at context dependency from the perspective of shared knowledge, even academic written discourse is highly contextualized. When comparing a conversation to a written academic article, Prince

(1981) found that while academic written texts carry more inferable entities (entities that demand logical reasoning by the addressees), spoken narratives carry more evoked entities (entities that are present in the situational context or that have been previously mentioned). Thus, Tannen and Prince have demonstrated that written discourse is a situated speech event too, given that its understanding depends highly on how much the community to which it is addressed shares background knowledge.

Nystrand (1987) confirmed the need for shared knowledge in written events by analyzing Olson's (1977) essay on the autonomous nature of written discourse, notes, and signs. Situational variables such as who, what, when, where, and how are critical to the meaning of those written text genres. Borrowing two of his examples, an EXIT sign would be meaningless if placed away from doors and a scholar's essay loses its value if not placed in the appropriate journal and field of inquiry. Take the case of notes written by children. They mean little for people who do not know them, their reasons and other situational variables (e.g. "I'm at the club. Keys are with me. Be back at 5."). What club and keys are these only the addressee and the addresser know. Simply put, as much as speech, written discourse is situated and interactional.

The notion that written and spoken language are in a continuum despite their different natures has been amplified by Chafe and Danielewicz (1987). Because the methodological steps I take in this dissertation modulate Chafe and Danielewicz's procedures (see Chapter VI), I treat their study in greater detail.

Unlike Chafe (1982) that analyzes only extremes of "spokenness" (conversations) and "writtenness" (academic papers), the Chafe and Danielewicz (1987) study treats four different genres in the spoken-written language scale: two spoken (conversations and

lectures) and two written (letters and academic papers). Data came from transcripts of 20 professors or graduate students whose natural conversations and lectures were taped and transcribed and who also provided letters written to family and friends and academic papers. Chafe (1982) used part of the same data.

To analyze the data, Chafe and Danielewicz (1987) examined quantitatively and qualitatively variety of vocabulary, level of vocabulary, clause construction, sentence construction, involvement, and detachment. To measure variety of vocabulary they computed the type/token ratio (the number of different words divided by the total number of words in a sample), and the mean number of occurrences of hedges and inexplicit third person references (*it*, *this* or *that*) per thousand words. To measure level of vocabulary, they computed the number of occurrences per thousand words of distinctly literary or distinctly colloquial vocabulary, and the occurrence of contractions.

To investigate clause construction they counted and averaged the number of words in an intonation unit (a spurt of spoken language with a single, coherent intonation contour followed by a pause and likely to be a single clause) and assumed that written language has a covert prosody, analogous to speech, indicated by punctuation. That is, for their written data, they counted the number of words between punctuation marks and treated them as if they were intonation units. That research procedure departs from previous research that has used the clause as the basic unit of analysis of written discourse. According to them, intonation units in spoken and written language have a cognitive basis, being the particular knowledge on which speakers are focusing their attention at the moment of speech or a reader at the moment of text comprehension. In their analysis of clause construction, Chafe and Danielewicz also computed the distribution of devices that

increase the size of intonation units, namely the number of occurrences per thousand words of prepositional phrases, prepositional phrase sequences, nominalizations, attributive adjectives and nouns, the use of "and" to conjoin two elements into a compound phrase, and the use of present and past participles.

To investigate sentence construction, the researchers measured the degree of coordination in the sample by counting the coordinating conjunctions ("and," "but" or "so") which were located at the beginnings of intonation units and expressed the measure as the percentage of intonation units which began in this way. They have also counted the number of running words in a sentence.

To measure involvement, the researchers counted the number of occurrences per thousand words of fillers such as "you know," first person pronouns, and temporal and locative adverbials. Finally, to investigate detachment, Chafe and Danielewicz computed the number of occurrences per thousand words of abstract subjects such as "this suggestion" or "educational setting," passives, and words that indicated probability such as "usually," and "probably."

Results indicate that academic lecturers and letters are mixed kinds of language that take advantage of some features typically recognized of writing and others of speech. While lecturers use literary vocabulary, extend their intonation units over 6 words, and show some degree of detachment (written-like speech), letters use intonation units of intermediate length (8.4 words), show great degree of involvement with the audience and with concrete reality, and a low degree of detachment (spoken-like writing). Academic papers and conversations represent the extremes of what writing and speech are. While academic papers show maximally varied vocabulary, few hedges, explicit references,

literary vocabulary with almost no colloquial items or contractions, conversations tend to hedge their lexical choices and be referentially inexplicit, showing colloquial words and phrases. Furthermore, academic papers create intonation units that are maximally long (mean average of 9.3 words), using prepositional phrases, participles, and nominalizations to expand the units. In addition, Chafe and Danielewicz indicate that sentences tend toward a “normal length” of 24 words. Unfortunately, they fail to mention the origin of this result or the discourse type to which it applies.

Chafe and Danielewicz (1987) attribute the linguistic differences and overlaps found across the four genres to cognitive and social reasons. If the use of linguistic devices is related to context, purpose, and subject matter, differences between speech and writing are easily overridden (a point similar to Prince’s, 1981; Tannen’s, 1982b; Nystrand’s, 1987). For example, the use of time and locative adverbials are contextually determined (see the case of notes and signs mentioned above). Some written language (letters for example) may be more spoken-like in this respect than any spoken language itself. It all depends on how much the writer and the reader share the same reality. In the case of vocabulary level, the constraints are imposed by judgments of appropriateness. Nothing keeps a lecturer from using some colloquial expressions or a conversationalist from using literary vocabulary.

Though Chafe and Danielewicz (1987) recognize writing as a speech event and the influence of context in the shaping of written language, they fail to give full importance to the interactional aspect of the writing process in their analysis. Constructions that they relate to speed of processing, such as intonation units, divide spoken and written language sharply. Apparently, because Chafe and Danielewicz see writing as a product, they free

writing from any memory constraint—“writing frees writers from the constraint which keeps down the size of spoken intonation units. They need not limit the production of language to what can be focused on at one time, but spend an indefinite amount of time constructing intonation units of any size” (p.96). Interesting enough, their own research shows that writers keep stretches of language between punctuation marks (units analogous to intonation units in speech) within the bounds of nine words in English. This is an observation of utmost importance, in my opinion. Chafe and Danielewicz briefly speculate that, perhaps, the fact has to do with the writer’s sensitivity to the task of the reader and that the stretches of language between punctuation marks may be units of comprehension. They not only overlook that writing is an interactional process, but also that crosslinguistic variation may also occur. Different languages pack different amounts of information in a word (see the case of Brazilian Portuguese and English, in Chapter V).

A closer look at the issue may show that speed of processing relates to length of those units in written language as much as to intonation units in speech. In addition, it may reveal that, according to the morpho-semantic characteristics of different languages, writers keep them within the bounds of a “normal length” because of the reader’s limited ability to shift the focus of attention. Chafe (1994) himself revises the issue when examining a different language (see under the heading “Texts as linguistic, social and cultural phenomena”). As Nystrand says “Writing is no less interactive in either principle or practice than speech (p.207) and “writers write on the premises of readers and readers read on the premises of writers, the result is communication and comprehension” (Rommetveit, cited in Nystrand, p. 207).

Like Tannen (1982b), Chafe and Danielewicz (1987), and Nystrand (1987), Biber (1988) shows that there is no simple two-way distinction between speech and writing. Rather, the two modes overlap because texts are multidimensional structures. Unlike previous researchers, he worked with computer-based text corpora, investigating 67 linguistic features in 481 texts of 23 different genres and used a Multi-feature Multi-dimensional analysis framework (1984, 1986) he developed. This framework included a computer identification and frequency count of features such as tense and aspect markers, place and time adverbials, nominalizations ending in *-tion*, *-ment*, *-ness*, *-ity*, passives, subordinating constructions, prepositional phrases, lexical specificity (type-token ratio), lexical classes, reduced forms, and coordination (clause initial *and*). Besides the overall distribution of the selected features in the data, the analysis included a principal factor analysis that clustered the co-occurring features into communicative functions that followed Biber's literature review.

Results indicate that accounts of texts may include at least six major dimensions of linguistic variation and each dimension indicates a different set of relations among texts. The dimensions of linguistic variation Biber defines are (1) involved versus informational production; (2) narrative versus non-narrative concerns; (3) explicit versus situation-dependent reference; (4) overt versus non-overt expression; (5) abstract versus non-abstract information; (6) on-line elaboration. Results also show that the frequency and distribution of linguistic features vary across text types, genres, and sub-genres. Therefore, differences between discourse modalities are also a function of text types and genres.

Furthermore, Biber (1988) shows that while text type characterizes discourse on the basis of form (narrative, descriptive, expository, argumentative) or features that are linguistics proper, genre characterizes discourse on the basis of use, purpose, and topic.

Of relevance to this dissertation are Biber's (1988) findings related to dimension one, involvement versus informational production. That dimension proved to be a significant predictor of genre differences. Annual reports and reportages such as those I will analyze emerged as typically high informational, with moderately-high level of explicitness and abstractness, and low level of informational elaboration. The high frequency of nouns and prepositions and the high type-token ratio of this genre in English have been interpreted reflecting a lack of concern with the reader's reality or non-involvement.

Another dimension that Biber (1988) reveals to be typical of the *report* genre is abstract versus non-abstract information. Annual reports obtained the second highest score on this dimension, showing frequent use of passives, agent deletions, and adverbial subordination. Because the report of the activity itself or the discourse topic are conceptual in nature, human agents are quite incidental, being largely unimportant in relation to activities. Therefore, they are always deleted or need to be inferred by the readers.

Other characteristics of texts in the institutional sub-genre (Biber, 1988) include frequent occurrence of phrasal coordination and nominalizations, infrequent occurrences of place and time adverbials and other adverbs; few verbs and few animate referents; presentation of action in the present tense; use of many attributive adjectives to provide descriptive details and elaborate on the nature of the nominal referent; and careful choice

of words. To summarize, according to Biber, the sub-genre I will analyze in this dissertation draws extensively on reference that is text internal or endophoric. That is, it is strictly informational.

Biber's (1988) description of relations among texts in English may serve as the basis for any functional text analysis provided that variations within major genres be observed. Most of all, he has shown that to realistically compare texts in different modes or across languages, one needs to address more than one dimension (see also Dantas-Whitney and Grabe, 1989) and keep genre and sub-genres as control variables. However, like Chafe and Danielewicz (1987), he overlooked the interactional aspect of writing as a process.

Other studies focused exclusively on written texts and their properties. Using Biber's (1984; 1986) Multifeature Multidimensional framework, Grabe (1987) investigated 150 expository texts of fifteen different types to characterize expository prose in English as a major text genre, a notion that contradicts Biber (1988). His findings point toward a division of expository prose into humanities texts, general informational texts, and two types of natural science texts. Among the 27 syntactic and 6 cohesion variables he analyzed were nominalizations, number of words per sentence, attributive adjectives, passives, relative clauses, locative adverbs, and lexical synonym/antonym. Confirming Biber (1988), some of those variables were more prevalent in certain text types. For instance, nominalizations, passives, and repetitions were more prevalent in natural science writing. That is also true of annual reports. This genre was closely linked to prose in academic natural science writing. Like Biber, Grabe's major contribution lies in future applications of this study for textual analysis, mainly contrastive analysis of expository prose across languages. Research should only examine texts that belong to the same sub-

type category to assure the validity of the study. In addition, crosslinguistic studies of that type would indicate how similar texts considered “expository” are in different languages and the textual dimensions that differ from English.

Texts as knowledge structures

Criticism of Halliday and Hasan (1976) led researchers such as Witte and Faigley (1980), de Beaugrande and Dressler (1981), Carrell (1982) and others to look for explanations of textuality outside the text or that which is linguistic per se. Witte and Faigley (1981), for example, show that “texture” has to do not only with overt linguistic properties like cohesion but also with real-world facts such as writers’ and readers’ perspectives, genre, knowledge of the topic and of other texts of the same genre, structuring of information (see also Tierney and Mosenthal, 1983; Almeida, 1984 on the interplay of cohesion and coherence). Such understanding has forced a departure from an exclusive functional approach to text analysis, bringing in an increasing interest in the role the mind plays in construing “texture.”

Psychological treatments of written discourse appeared in the 1970s with van Dijk’s (1972), Stein and Glen’s (1979) and Mandler and Johnson’s text grammars (1977), Kintsch’s (1974) propositional system, Kintsch and van Dijk’s (1978) macrostructure analysis, and Meyer’s (1975) rhetorical predicates of expository prose. Those studies view texts as reflections of psychological processes involved in producing and comprehending language. Among other things they show that textual macrostructures or top-level rhetorical organization plays an important role in comprehension (van Dijk; Kintsch and van Dijk; Meyer). Readers tend to focus on and remember information at higher levels. While adherence to canonical story parts strongly predicts comprehension

of narratives (Stein and Glen; Mandler and Johnson), semantic structure (Kintsch and van Dijk; Meyer) or macropropositions reveal the semantic acceptability of a text, vis-à-vis previous propositions and the comprehenders' past experiences with texts (see Chapter IV for further detail on how memory for texts relates to information structuring).

Accounts of structural hierarchy in texts and its influence on comprehension and recall found support in Schank and Abelson's (1977) notion of scripts, according to which our expectations are conceptual rather than lexical. Namely, the mention of the word "restaurant" primes a set of actions in a temporal sequence, from the goal—to obtain food—to the ordering, to the payment of the bill. Other possible names for those higher-level knowledge structures or ideational scaffoldings that predict text comprehension are schemata as defined by Rumelhart (1977, 1980) and Carrell (1983), frames of expectation as defined by Tannen (1979) or mental models (Johnson-Laird, 1983). Analysis of discourse that takes such understanding into account integrates linguistic information present in the text with knowledge structures that pre-exist in the minds of writers and readers: "When we read a text, we match it against a network of experiences of texts we regard as comparable situations and contexts" (Enkvist, 1987, p 38). Chafe (1994) puts it this way: "A fundamentally important property of human consciousness is its ability to focus, not just on the immediate environment, but also on remembered or imagined experience" (p.210) where imagined experience dwells in the domain of scripts, schemas and the like.

This line of research has contributed to the understanding of coherence as a process that takes place in the mind, rather than as an intrinsic linguistic trait that only appears in the text (Charolle, 1983; Samet and Schank, 1984; Givón, 1995). As Connor (1996) says,

coherent texts simply make sense to readers. In addition, this research has boosted crosscultural studies that have dealt with perceptions of writing quality of ESL students (Almeida, 1984; Connor, 1984; Saliés, 1995b). If one or more categories of expectation fails to match the reader's own conventionalized frame, texts are perceived as less coherent (see the contrastive rhetoric heading for details on the role of knowledge structures in text comprehension and production in the field of Second Language Acquisition).

However, though those psychological accounts of texts helped us to understand the role of information and mental structuring in text comprehension, they have overlooked how linguistic information present in the text guides the process of building coherence (Grabe, 1996), and function as traffic signs or tracking systems toward comprehension (Givón, 1995). According to Givón, grammatical structures (lexical and syntactic) act as basic processing directions for language comprehension, pointing forward (cataphoric signalling) or backwards (anaphoric signalling) to maintain a topic, reinstate it, or present a competitor. In his view, the syntactic systems that help the reader to construct a mental model of the text are references (pronominals, repetition, structures that reinstate referents, definiteness), temporality, aspect, mood, location, scripts, conjunctive and disjunctive relations, and grammatical relations such as transitivity. His view is similar to Halliday and Hasan's (1976; 1989), Celce-Murcia's (1990), Tannen's (1989), Chafe's (1994), Haiman and Thompson's (1984) who see grammar as a processing mechanism for constructing discourse. If you do not master bottom-up strategies, you can hardly construe a text mental model (I direct the reader to Chapter IV where I discuss Givón in detail).

Of relevance to this study is Givón's (1995) notion about the relation between lexical/grammatical information in the text and processing speed. For him, while lexical categories unfold slow but strong processing instructions, grammatical devices provide fast but weak processing directions. That is, grammar provides a first and incomplete picture of the intended message and is automatically processed. The lexical categories sharpen that image slowly, providing specific and strong activation links between items in a text. As Grabe (1996) mentions, the grammar provides early activation of related sets of information that are later strengthened or suppressed through specific lexical information. In Chapter IV, I shall go back to this issue in detail.

Building on the understanding of texts as a set of linguistic devices that guide the readers to construct a mental model, several studies have emphasized the interactional aspect of the writing process or else, texts as communicative acts. Those accounts are the concern of the next sub-heading.

Texts as situated interaction

Written accounts of texts as situated interaction look at discourse as a give and take between writers and readers within a psycho-social context, explaining the referential meanings of structures in view of that context. Texts take the form they do because of assumptions writers make about their readers' reception of language. Mostly, such type of analysis accounts for sequential expectations that emerge from situating morphological, syntactic, and rhetorical cues in the cognitive and social universe of readers. To do that, it assumes that readers will make inferences and comprehend based on what is private (their cognitive capacities) and what is social (knowledge that they share, including the linguistic system).

Major proponents of interactional mechanisms in discourse are de Beaugrande (1980) and de Beaugrande and Dressler (1981). De Beaugrande, specially, argues that a text cannot be analyzed apart from the procedures humans use to produce and comprehend them. For him, textuality lies in the psychological and social activities language users perform and presupposes knowledge of the world, of whole-part schemas, of paronymy and hyponymy, among other things.

Other proponents of texts as situated interaction (Chafe, 1994; Lautamatti, 1987; Prince, 1981, 1992; Brown and Yule, 1983; Connor, 1987 and others) call on how speech event participants use concepts such as “topic” and “comment” to construe meaning. The asymmetry between topic and comment marks information that writers believe to be somehow known to readers (old information, the topic) and that which they believe to be unknown (new information, the comment). It has been labeled the given-new contract (Haviland and Clark, 1974; Clark and Clark, 1977), and interpreted and sub-divided in slightly different ways by researchers in the field.

While Clark and Clark (1977) interpret “given information” as information that is merely identifiable in discourse or that has an antecedent in memory, Chafe (1987) narrows the use of the term. For Chafe, “given” information refers to events, participants, or ideas the addresser assumes to be in the mind of the addressee at the time of communication. Actually, he connects time and consciousness to the production and comprehension of discourse, establishing *givenness* as a transient aspect of it. According to him, information can have at least three states within the mind: active or immediately related to the reality of an individual at the moment of communication; semi-active or accessible from the previous context; and inactive or new at a specific point of discourse.

Semi-active information, though previously mentioned in preceding discourse, has been displaced in time by other entities or events, being peripheral to consciousness. Therefore, its re-activation involves a cost that may be operationalized as the amount of time it takes addressees to retrieve that information from the knowledge base and re-instate it in their immediate reality. In his view, the best way to understand the given-accessible-new distinction is by calling on notions such as cognitive cost or in his words, “activation cost” (1995, p.73).

The interaction between activation cost and the status of information in discourse (given or new) was first addressed by Haviland and Clark (1974) in reaction time experiments. Haviland and Clark prove that whenever two sentences have no direct linkage, requiring an extra inferential step, readers take longer to push the button or to make the bridge between what was actually known and what was being treated as given.

Indeed, Chafe’s (1994) investigations led him to conclude that new information is not an idea or event that was unknown to a language user; rather, it is information that was inactive up to a point and that then was newly activated. By extension, he characterizes given information as information that was already active at a specific point in time. Therefore, “given information is least costly in the transition from T1 (the first point in the time-line) to T2 (a second point in time-line) because it was already active at T1. Accessible information is somewhat more costly, and new information is the most costly of all, presumably because more mental effort is involved in converting an idea from the inactive to the active state”(p.73). Note that only content words (noun phrases, pronouns, verbs and adjectives), as opposed to function words, are associated with

activation cost. Content words constitute referents, events and states or simply, ideas.

This study will treat such structures.

Chafe (1987, 1994) emphasizes that what happens in people's minds when they manipulate given and new information follows a continuum. That is, the "active," "inactive," and "semi-active" states are not categorical; instead, they dynamically shape discourse. While events and ideas are constantly being replaced by other events and ideas, remaining in the consciousness of addressees for not more than a second or two, referents remain active for longer periods, sometimes across intonation units (units that carry a single idea; see my discussion of Chafe and Danielewicz, 1987, above). According to Chafe, it is that flow from active to semi-active to inactive and back to active states that reflects human mental experience when producing and comprehending a text.

The notion of semiactive consciousness introduced by Chafe (1987, 1994) is especially relevant because it allows a refined understanding of how language users place active information within large frames of discourse that require a larger capacity to store and combine ideas. As Chafe (1994) says, sets of information that are active at a given moment (one information unit) combine to form larger unified sequences (sentences) which combine to form larger chunks of semantically and syntactically unified discourse (topics). Those larger chunks do not fit the capacity of focal consciousness. Therefore, it is semiactive consciousness that holds a topic, a psychological notion Chafe (1994) defines as an aggregate of coherently related events, states, and referents. Once a topic becomes semiactive, it may be sustained through elicitation or narration in discourse. The length of time a particular topic is held in the semiactive state varies according to the addresser. On the other hand, the length of time a particular topic is held in the active state at a given

time relates to the capacity of focal consciousness. According to Chafe, that capacity in spoken language consists of one intonation unit or one new idea at a time and reflects a limitation on the speaker's own processing capacity, but also the speaker's own awareness of a parallel limitation in the mind of the listener.

Though Chafe's (1974, 1976, 1987, 1994) understanding of *givenness* cannot be overtly observed, it emerges from analysis of real language as evidence to processes that take place in the mind. First he observes natural data, counting specific features and relating structure to its context of use. Then, he makes theoretical generalizations (introspections) that explain information flow in view of the cognitive universe of the participants of a given speech event.

Another aspect of Chafe's methodological procedures that I would like to highlight (Chafe and Danielewicz, 1987; Chafe, 1988) relates to how the boundaries of intonation units in written language were established without the help of prosody. Chafe and Danielewicz (see the review above) assume a relation between spoken intonation units and segments of language bounded by punctuation marks (labeled written intonation units in that study). Chafe (1988) further discusses that relation, and concludes that despite inconsistencies (writers vary in their ability to use punctuation effectively, styles change), there is a tendency to punctuate in the way speakers divide intonation units. Stretches of language between punctuation marks provide some approximation to spoken intonation units. That is, both intonation units and segments of language bounded by punctuation marks apparently stick to the "one new idea constraint" (Chafe, 1994).

Therefore, it is worth examining punctuation units "as if they did reflect intonation units" (Chafe, 1994, p.291). As Chafe (1988) himself stresses, the validity of this

methodological procedure deserves further investigation. Other crosslinguistic studies shall throw light on the role of stretches of language between punctuation marks in written discourse and their relation to spoken intonation units.

Other accounts of written discourse that worked with information structure contributed slightly different understandings of the given-new contract. Prince (1981) has sub-divided the “new” category into two sub-categories—brand new and unused. According to her analysis of conversational language, brand new is that information which is typically new or that has just been introduced into discourse, being normally expressed by indefinite expressions such as the indefinite article *a*. Unused is that information the addresser assumes to be in the background knowledge of the addressee. She has also added a third category called “inferrable” or information that can be inferred from the text. For the “given” category, she indicates information that has been evoked by the text or the situation.

Brown and Yule (1983) have expanded Prince’s 1981 taxonomy further dividing the “given” category in current (what Chafe labels under the focus of attention or being directly perceived, active) and displaced (that cannot be directly perceived but remembered, semi-active).

Prince (1992) herself revised her 1981 taxonomy of information status. In her study of the ZPG letter (Zero Population Growth 1985 Stress Test), she investigates the relation between subjecthood, definiteness and introduces a distinction between discourse and hearer old or new entities. ZPG is a 1985 survey of “how population-linked pressures affect U.S. cities” (p. ix), and the letter is an appeal for money signed by its director. According to Prince’s analysis of the letter, discourse entities may be old or new with

respect to the hearer or old or new with respect to the discourse. It follows that if an entity is hearer-new, it is discourse brand-new in the old terminology. If it is discourse-new unused, it is hearer-old. The inferrables are hearer-new and discourse-new if hearers' background knowledge or some trigger entity in discourse has not made it hearer-old and/or discourse-old. Like in the 1981 study, the category "Containing Inferrables" carries within it the entity that triggers the inference. A containing inferrable is normally syntactic complex (e.g. nominalizations), containing a NP and a subordinate clause.

The studies that addressed information status and their relations to production and comprehension of referring expression in discourse (Chafe, 1987; Brown and Yule, 1983; Prince, 1992) assume high interaction between what writers believe readers know, may know or do not yet know. While they studied mainly the effects of those notions in relation to referring expressions, Erteschik and Lappin (1979) examined the information status of the subordinate clause, finding that it is the position of the clause, sentence-initial or sentence-final, that determines its status as foregrounded or backgrounded information (active or semi-active in Chafe's terminology). If sentence-final, subordinate clauses may have a foregrounding function. Sentence-final position is a parameter for establishing dominance or activeness in consciousness.

Research on non-native accent in foreign and second language teaching (Erlach, 1988; Saliés, 1995b), further illustrates the interaction between information status, production, and comprehension. Because language learners use syntax to foreground and background elements in ways that differ from native speakers, they impact coherence and lead to negative judgements of their writing. Saliés, for example, case-studied compositions written by two Malay students. Her analysis demonstrates that subjects used

Y-movement constructions in ways that violated the raters' expectations. This led to discrepancies in the ratings and low inter-rater reliability. In English, focus of assertion is topical, and involves the speakers' evaluation of the hearer's attitude toward the information. Therefore, Y-movement constructions can only be used if the expectation has been set up in the immediate previous text. Native English speakers do not expect those constructions to introduce a topic or be discourse-new. When this is the case, it is more difficult for readers to identify the main discourse topics and build coherence.

The interaction of what writers believe readers know, may know or do not yet know has also been investigated crosslinguistically. Woodley (1986) analyzed texts written by French native speakers, English native speakers, and students of French as a foreign language. English texts in this study differed from those in French in the way they used clefts. Most native speakers of French used clefts to represent information they took to be discourse-old or given. However, that information was definitely not known to the readers, but discourse-new. On the other hand, English native speakers used them to stress focus type. That is, their clefts represented information they assumed readers knew or could deduce from context. Readers were left with a great deal of inferring to do.

Not only studies that have focused on the linear organization of discourse and its relation to the mind have used an interactional approach. Connor (1987) has analyzed cross-cultural differences in the patterns of argumentative texts of ten natives of England, Finland, Germany, and the United States within an interactional and interdisciplinary approach to text analysis. She posits that texts are the embodiment of the writer's knowledge about schema-based production or the realization of intentions through linguistic devices. Using a three-level analytic system--argumentation as a situation +

problem + solution + evaluation structure; argumentation as a succession of speech acts; and argumentation as a direct appeal to an audience—she randomly selected ten compositions from each of the four countries, had them rated for overall impression by three independent raters from Purdue University, and correlated the scores with a measure of audience awareness proposed by Delia, Kline, and Burleson (1979, cited in Connor). Results show a slight difference in the average holistic scores between the different language groups. The highly-rated compositions followed the pattern situation + problem + solution + evaluation. Even the low-ranking English and U.S. essays largely followed the structure. With regards to the second level of analysis, the speech-act sequence, high-rated essays asserted a claim, justified and induced it, including directive inductions in the solution section. Last, the judged level of audience awareness correlated highly with the overall quality of the compositions. That is, if we knew a student scored high on the audience awareness scale, we could predict his overall quality score with 46 percent accuracy. Though Connor failed to find cultural variations with regard to audience awareness, she found some as related to structure: the problem-solution schema was not used as consistently in the Finnish and German compositions as it was in the English and U.S. compositions. However, the importance of audience awareness was underscored as a predictor of successful writing, showing that writers have specific reasons for selecting ideas as well as for deciding how to present them. The ability to represent the characteristics and perspectives of a potential reader played a central role in the construction of the persuasive messages.

Lautamatti (1987) has reinforced Connor (1987), asserting that readers have expectations about the way written texts are structured. One of the things readers expect

is that the sequences making up a piece of text be somehow related to the main idea or the discourse topic. In this paper, she examines aspects of topical development by comparing an original English text (*Language and Community* by Anne and Peter Doughty, 1974) to its simplified versions. The simplified versions were provided by Alan Davies and Gillian Brown, from the Department of Linguistics, University of Edinburgh, with the purpose of making the text more comprehensible. Specifically, she investigates the role of subjects in topical development, focusing on mood subjects of the main clause (notional subjects that are in subject position); topical subjects (notional subjects that are in subject position and represent what the clause is about); and initial sentence elements (modality markers and connectors). The four simplified text versions, as opposed to the original text, favored constructions where the topical subject is the mood subject of the main clause, or, if that was not the case, structures where the initial main clause serves as a modality marker or connector and the topical subject follows immediately in a sub-clause. For example, one of the simplified versions reads (p.111):

All healthy, **new born babies**...share two characteristics
The first characteristic which all human babies share, is that **they** are completely helpless.
 The only thing **they** can do to persuade someone to look after them is to cry...

while the original text reads (Lautamatti, 1987, p.92):

When a **human infant** is born into any community in any part of the world, **it** has two things in common with any other infant...
Firstly, an most obviously, **new born children** are completely helpless.
 Apart from a powerful capacity to draw attention to their helplessness by using sound, there is nothing **the new born child** can do to ensure his own survival.

Results also indicate that some types of topical progression are intuitively felt to be more readable than others. That is the case of main clauses with syntactically prominent

mood subjects (= *all healthy new born babies*). Likewise, parallel topic progression (the topical subjects of one of the simplified versions were = *Baby / it / it / baby / humans / baby / humans / babies*), contrary to sequential progression, seems to increase readability. The simplified versions showed a decrease in the number of sub-topics per sentence and an increase in the use of parallel progression. That is, the topical subjects of successive sentences kept the same referent more frequently than they gave rise to new sub-topics.

Finally, from a theoretical perspective, Grabe and Kaplan (1996) explore an ethnography of writing and propose a model of writing as communicative language use. They argue that when people write, they tell and transform knowledge in a social context that consists, among other things, of a setting, participants and interactions (audience, Gricean Maxims, politeness, power, face, functional purposes), and of on-line processing assembly of the pieces and bits within verbal working memory.

In brief, the studies reviewed reveal that like spoken language, written language is used by writers in real contexts to achieve specific goals. To explain the structures present in it, analysts call on cognitive notions such as activation cost, constraints on information flow, coherence, topic development, and on addressers' intentions and addressees' expectations. Meaning thus dwells in what is shared, be it conceptual structures of text organization or grammatical and lexical information present in the text, and on what is private (e.g. encyclopedic knowledge).

Other researchers have enriched our understanding of written discourse by situating it in a slightly different context: The context of social norms of communication. For those researchers, participants have expectations because of norms of communication that they share, not because of cognitive capacities or patterns of information flow. These norms

guide the give and take between participants and constrain the choice of structures. That is the focus of the following exposition.

Texts as pragmatic inferences

Studies that situate discourse within the context of communication concentrate on what writers and readers know about social institutions, others' needs, the nature of human rationality, and how that knowledge guides the use of language (see Schrifin, 1995). According to these studies, communication derives from the interaction between the text, its contexts (social, cognitive and linguistic), and addressers' and addressees' mutual knowledge about social communication norms.

The originator of this view of discourse is Grice (1975). Though he has never applied his views to the analysis of natural discourse, he provided a set of principles that indirectly contributes to analyses of written discourse. According to him, every act of communication is guided by norms of conduct or maxims that combine to form the context of the speech event and to allow addressees to make inferences about the intended meaning. These maxims are called the cooperative principle (CP) and say that participants should offer just the information required, at the right time, and following the purpose and norms of the exchange.

Specifically, the Cooperative Principle includes the maxims of quantity (offer just the information required for the purpose of the exchange, not more nor less); quality (offer just information that you believe to be true and that you can prove); relation (offer just information that is relevant); and manner (avoid obscurity of expression, ambiguity, and wordiness; be brief and orderly).

The operationalization of those maxims rests upon the concept of implicature. That is, inferences users make about the intended meaning, based on semantic information present in the text and on shared knowledge. Whatever inferences of a writer's meaning a reader makes, they have to be based on the linguistic information, the context (linguistic or situational), background knowledge, and the fact that both reader and writer share those notions.

A Gricean analysis of written discourse examines, then, how meaning depends on shared information and the assumptions on which a writer relies to construe communication. This understanding is similar, in some respects, to Grabe's (1996) and Givón's (1995) argument that the linguistic code guides interpretation. However, the focus is on "individual, intention-based meaning that could supplement the logical, propositional and conventional meanings representable through a linguistic code" (Schriffin, 1995, p. 408).

Schriffin (1995), using natural spoken data, demonstrates how the Gricean maxims of quantity and relevance explain the sequential alternation between referring terms in a narrative, particularly the use of indefinite/definite, and explicit/inexplicit referring terms. The textual and contextual information provided in initial position (first mention) in her data served as background against which to judge how much information was appropriate in a next position (next mention). It also served as a means to judge the relevance of that information for the second position. Thus, first mentions which were indefinite and explicit presupposed less about the ability of addressees to identify a referent than next mentions which were likely to be definite and less explicit than first-mentions. However, when a referent could be interpreted as relevant to already accessible information,

subsequent definites would also occur. She concludes that explicit next mentions can be viewed as a violation of the quantity maxim designed to show relevance.

While the maxim of quantity helps to guide addressees toward information that can provide clues about the identity of a referent, the maxim of relation leads the addressee to search for the relevance of a particular referent in the contexts of a text (cognitive, social, and linguistic). Definiteness becomes then that which the addresser expects the addressee to know and explicitness that which is textually presented to enable the addressee to identify an intended meaning. It is by means of the contexts of a text that addressees interpret or infer the addresser's meaning in different speech acts and it is by means of the operation of Grice's cooperative principle in one part of discourse that users define options in another.

Similarly, Jones (1983) examines pragmatic aspects of English written nonfiction texts, showing how lexical and grammatical structures reflect authors' assumptions about readers' knowledge. Specifically, he focuses on first mention references, technical terms, and author comments, concluding, like Schriffin (1995), that those categories have to do with author assumptions and the notions of foregrounding and frame. If authors presuppose that readers will be able to locate a referent in their foregrounded frame (under the focus of attention), they use definite expressions to guide the reader. In addition, Jones cites Becker (1980) to say that part of the context of any text is all previous texts in a particular culture, specially those of the same genre. A given text will simply be an instantiation of that knowledge.

Given the role relevance plays in guiding users' choices of forms and consequently, reader's focus of attention, Sperber and Wilson (1986; 1995) further developed the

Gricean notion of relevance into what became known as the Relevance Theory. Such theory says that to communicate is to imply that the information communicated is relevant. Basically, the Relevance Theory has solved the problem of how to select a context for interpretation. Sperber and Wilson explain that “The actual context used in interpreting an utterance is constrained by the organisation of the individual’s encyclopaedic memory, his perceptual and other cognitive abilities, and the mental activity in which he is engaged” (p.138). The most recently processed information is more accessible (see my review of Chafe, 1994) and it will combine with the information derived from it to yield the context of interpretation.

Besides, given that different users require different amounts of contextual cues to recover the relevant information from a text, the Relevance Theory also says that for optimal communication writers should make sure the contextual effects are maximized so as to make retrieval of meaning in a particular situation as easy as possible. In other words, writers should make sure their messages do not put readers through an overly difficult processing effort. According to Sperber and Wilson (1995), any increase in the processing effort is a risk to comprehension because it detracts readers from overall relevance—“Relevance is itself defined in terms of positive cognitive effects” (p.266). If the information communicated is relevant, addressees may be sure that it implies the smallest processing cost. That is, the information conveyed combines with the context to yield contextual implications by strengthening or contradicting existing assumptions (this understanding is similar to Givón’s argument, 1995, and fits Kintsch’s model of comprehension, 1995) or yet create new assumptions. If the information has no contextual effects, it is not relevant.

Actually, the Relevance Theory reinforces, from the perspective of communication studies, the argument that textual features are instructions toward comprehension (Grabe, 1996; Givón, 1995; Kintsch, 1995), and that, as Givón mentions, they generate different processing speeds. An analysis within the framework of the Relevance Theory assumes that written texts carry strong implicatures, or contextual effects, that leave a small margin for alternative interpretations. Those contextual effects may be of three types: contextual implications (hypotheses that emerge from the physical and psychological environments); strengthening an existing assumptions (further evidence to a given implication); and contradicting and eliminating existing assumptions. Those effects determine the amount of processing effort.

Blass (1990) applies Sperber and Wilson's (1986; 1995) Relevance Theory to analyze the Sissala language. She shows that addressers use particular linguistic phenomena to achieve pragmatic effects. In Sissala, for example, the particle *dé* is used to minimize processing effort in that it guides the addressee to the most important information in the utterance. She argues that what is crucial to discourse comprehension are relevance relations because, as Sperber and Wilson discuss, the majority of contextual assumptions derive from memory and perception. In other words, contextual assumptions derive from the cognitive environment that addressers and addressees share. The physical environment only plays an indirect role in the retrieval of meaning by affecting the individual's cognitive environment. In other words, she argues for a psycho-pragmatic account of discourse.

Similar to Sperber and Wilson, Blass (1990) views relevance as a psychological constraint that emerges from contextual effects and processing effort. Concretely, that

could be translated into the idea that relevance is directly related to comprehending in the least time possible. If the amount of processing effort decreases, relevance increases. So, the accessibility of contextual assumptions relates to cost of processing and is key to comprehension. Textual structures are themselves determined by considerations of relevance.

In conclusion, users textualize information the way they do because of the Cooperative Principle and of the Principle of Relevance. The criterion of consistency with the principle of relevance assures analysts that linguistic devices present in a text aim at creating the optimal context for a single, low cost interpretation. Analyses of texts as sequences of pragmatic inferences assume information in the text to be maximally relevant for optimal processing.

Texts as linguistic, social, and cultural phenomena

Research on writing in a second language and contrastive rhetoric has shown that writers, in different languages, produce rhetorically distinct texts because literacy skills are not only shaped socially, culturally, and linguistically, but are learned (Grabe and Kaplan, 1989; Scribner and Cole, 1981; Heath, 1983).

Kaplan (1966) pioneered the idea that rhetoric “is not universal” (1966, p.2), but evolves out of culture in his analysis of English, Romance, Oriental, and Semitic expository prose. Since then, numerous studies have been examining rhetorical and other discursive features across languages .

Ostler (1987) has contrasted T-units and discourse blocs of twenty-one expository essays written in English by Saudi Arabian students and by native English speakers. While English subjects used two sub-divisions for each discourse unit at the most, Arabic-

speakers used three (77% of the papers), four (19%) or even five sub-divisions (4%), developing the topic in parallels. In addition, one-fourth of the sample ended the papers with some type of formulaic or proverbial statement.

Silva (1992) and Saliés (1995c; 1995d) have investigated students' perceptions of differences between writing in L1 and L2. Chinese subjects revealed that Chinese paragraphs are more inductive (Silva), with statements typically supported by citations of historical events. Malay (Saliés, 1995c) and Japanese (Silva) native speakers reported valuing indirectness in their writing and avoiding written conclusions. They prefer to be vague whenever native speakers of English value a more deductive and direct approach.

Hinds (1983, 1984) discusses the rhetorical pattern of expository prose in English and Japanese. His analysis indicates that Japanese speakers tend to develop sub-themes that do not bear a direct relation to the major topic, in the same fashion that Arabic (Kaplan, 1966; Attari, 1984; Ostler, 1987), Chinese (McKay, 1989; Silva, 1992) and Romance languages (1966) speakers do. A similar structure is identified in the literature for Korean academic prose (Kaplan, 1972; Eggington, 1987). Instead of developing issues in terms of what they are, Korean writers draw on what they are not, showing the subject from a variety of tangential views.

Clyne (1981, 1987, 1991) reports on the general non-linear structure of German as opposed to English. He notes that in German content takes a bigger role than formal style and that digressions and repetitions are much more tolerated than in English.

Linguistically, German writers' preference for a more elaborated style produce frequent use of nominalizations, syntactic complexity, agentless passives, and overloaded phrases. Because many German writers writing in English retain their German rhetorical

preferences, American scholars see that type of writing as chaotic and disorganized. Clyne (1981) exemplifies his point citing Dittmar's *Soziolinguistik*, a masterwork in sociolinguistics that got a negative review in America because of its pretentious style and bad organization of ideas. The translator failed to translate the discourse structure to English and the German non-linear structure in English left American readers lost, in search for meaning. Text organization, thus, has communicative value. The way in which a writer organizes information and arranges it on the page can be as important to effective communication as factual details themselves.

The notion that different cultures favor different rhetorical patterns is particularly valid given the evidence that languages differ in the way they identify and develop discourse topics (Mc Kay, 1989; Scarcella, 1983; Hu, Brown, and Brown, 1982; and Silva, 1992). In Scarcella's comparison of English with Spanish spoken discourse, native speakers of Spanish used far more personal topics when in informal conversations than English speakers. Hu, Brown, and Brown revealed that Chinese EFLers, as compared to Australians, preferred to focus on the importance of education for the nation as a whole in an essay to persuade a brother who does not work hard at school. Australian students stayed at the individual level. McKay compared essays on the same topic written by a group of students of various ethnic backgrounds living in San Francisco and a group of Chinese students studying English in the People's Republic of China. While Chinese EFLers developed the topic in a strikingly common way, closing their essays with a moral lesson, ESLers in the United States used a wide range of strategies, drawing no moral lesson in the conclusion. Simply put, depending on the culture, writers choose different

ways to express an idea or hammer a point. While Chinese and Arabic speakers prefer to draw on ancient history, parables or sayings, English speakers prefer facts.

The influence of social experience on writing has been primarily documented through research that dealt with readers' expectations (Hinds, 1987; Yoshikawa, 1978; Leap, 1989; Silva, 1992; see also studies on memory in Chapter IV). Hinds contrasting English to Japanese expository prose affirms that the degree of involvement a reader will have with a text depends on the language. In reader-responsible languages such as Japanese, readers supply some significant portion of the propositional structure, being primarily responsible for effective communication. What is not verbalized enjoys the status of truth and the reader is supposed to know it by intuition (Yoshikawa, 1978 cited by Hinds). It follows that Japanese writers do not offer thorough explanations of their views, but hints. On the other hand, in writer-responsible systems such as English, writers supply most of the propositional structures and that kind of appeal would be exactly what the audience would expect. If communication fails, it is because the writer has not been specific or clear enough. As emphasized by Hinds, this view of communication has permeated the culture of English speakers (Havelock, 1963, 1976; Chafe, 1982).

Leap (1989), Saliés (1995c; 1995d), Silva (1992) and McKay (1989) have also documented the influence of socially expected behavior on written language. Leap, in an ethnographic study with Ute English speakers, discussed how Ute English writers are selective in their use of description and narration, suggesting details instead of explicitly presenting them. In other words, the discourse framework and presupposition pool is assumed to be reader conscious because they are part of a shared social code. This finding is similar to what Saliés (1995c) found in a study of Malay students. For example, the

Malay students revealed to be impolite to include information that is socially shared in written texts. The writer simply outlines the message, demanding the active engagement of the reader to fill in the gaps. In French, however, Silva (1992) found a reference for more elaboration. One of Silva's subjects reported on how written French values style and an elaborated rhetoric, full of metaphors, expressive sentences and very complex phrasal structures. Though the French style renders explicitness it also yields digressions from the central topic.

Reid (1988), Montano-Harmon (1988; 1991), Lux (1991), Lux and Grabe (1991), and Reppen and Grabe (1993) found similar evidence for texts written in Spanish. Spanish writers (adults, adolescents or children) favor coordinating structures with lengthier and more elaborated sentences whether writing in Spanish or English. They also tend to use loose association of clauses (Reid and Montano-Harmon), more synonyms and additive conjunctions, favoring flowery language.

McKay (1989), besides investigating rhetorical structure (see above), showed that social experience predicts topic development. Her ESL students who were living in San Francisco drew on their social experience in the United States to describe a bus stop scene and the consequences of heavy rain. They not only described the rainfall in terms of weather forecast, but provided excuses for riding a bus instead of driving. In addition, these subjects showed concern with time pressure, a behavior hardly common in the social environments of their countries of origin.

Contrastive studies have also focused on the linguistic features of L2 written texts responsible for a discourse accent (Hu, Brown, and Brown, 1982; Johns, 1984; Connor, 1987). Hu et al., besides investigating structure (see above), examined the overall

distribution of cohesive devices in the English writing of their Chinese and native English speaking Australian students. They found a Chinese preference for conjunctions and an Australian preference for lexical cohesion. Johns replicated the Hu et al. study at the Shanghai Foreign Language Institute. Chinese writers in that study also preferred to use conjuncts to build cohesion, specially additives, when compared to NSs of English.

Investigations of ESL writing and contrastive rhetoric studies clearly demonstrate that cultures favor different ways with words. Similar to other behaviors in our daily lives such as greetings or manners at the table, written discourse has a culturally accepted paradigm or a paradigm that is readily accessible in users' cognitive frameworks. In Kaplan's (1987) words, if one intends to produce texts to be read by village women in Southeast Asia, the texts should reflect schemata that the women carry to facilitate information storage and retrieval (p.20).

Why is it so? For the same reason that information structuring influences comprehension. When readers encounter a text, they see it as one instantiation of the text category in a given genre. Their presupposition pools interact with the linguistic information that the writer judged relevant and included in the text (see my discussion of knowledge structures above) to build coherence. In writing pedagogy, we label that issue as "identification of audience expectations"

From a slightly different slant, Chafe (1994), Damron (1997), and Biber (1995) have also contributed to our understanding of how and why languages organize texts in different ways. Chafe's investigation of the North American Indian language, Seneca, demonstrates how the capacity of focal consciousness relates to the way different languages pack information in a word. Compared to English, Seneca words contain more

information, a fact that seems to yield spoken intonation units half as long (normally of two words) as those of English (normally of four-six words), measured in terms of words per intonation unit. Also, Seneca's use of pronominal prefixes, though similar to English's use of personal pronouns, does not derive from activation cost, since the prefixes do not typically express given referents. Rather, pronominal prefixes in Seneca function as core participants in events and states, and distinguish between agents and patients. They mark the distinction between subject and object. However, spoken Seneca is similar to spoken English in that most intonation units adhere to the "one new idea" constraint or the ability to keep only one new idea in each spoken intonation unit under the focus of attention. This fact suggests a cognitive constraint on how much information (perceptual-like events and states) can be fully active in the mind at one time. Languages that pack more information into a word such as Seneca should then show fewer words per intonation unit than English.

Surprisingly, Damron (1997) reveals quite the opposite. In her investigation of spoken Urdu and Pakistani English, Chafe's (1994) one new idea constraint per intonation unit does not hold. Urdu, though a highly inflected language, yields multi-clausal intonation units that contain many independent clauses. Some intonation units have up to three new ideas. Furthermore, Damron theorizes about the role of multi-clausal intonation units and processing constraints they may bring to bear. She concludes that the multi-clausal units signal contrasting topicalization and proposes an inferential model or schema that allows listeners to process information with a minimum effort. Because of the limitations of this study (it investigates two conversations), other studies should test the proposed inferential model for other genres and most of all, for other languages.

Through a slightly different vein, this dissertation contributes to the discussion of intonation unit length and the amount of information they carry. It investigates length of attention units in Brazilian Portuguese and English. Like Seneca and Urdu, Brazilian Portuguese (BP) normally carries more information in its words than English mainly because BP is an inflected language. Therefore, it may be that its written intonation units will be shorter than those of English.

Another contribution to our knowledge of how linguistic features contrast crosslinguistically is Biber (1995). This study extends Biber's 1988 study and applies the Multi-Dimensional (MD) analytical framework (1984; 1986; see my review of Biber 1988 for details about the MD model) to three other languages—Nukulaelae Tuvaluan, Korean, and Somali. Specifically, it compares the patterns of spoken and written register variation across those languages to the synchronic MD studies of English (1988). In addition, Biber (1995) provides a sociocultural description of the four language situations; a synchronic and diachronic analysis of the similarities and differences among registers; and the text types that are well defined linguistically in each case. Some of the registers analyzed are reportages, editorials, novels, conversation, television documentaries. The linguistic features include tense and aspect markers, place and time adverbials, nominalizations ending in *-tion*, *-ment*, *-ness*, *-ity*, subordination features, and prepositional phrases. Details about the methodology may be found in my review of Biber (1988) that appears earlier in this chapter. Among other things, Biber's (1995) results indicate that the linguistic features that define each dimension (e.g. involved versus informational production; narrative versus non-narrative concerns; explicit versus situation-dependent reference; overt versus non-overt expression; abstract versus non abstract information) are

similar across the four languages. There are aspects of form-function that seem to relate to human communication in general. Those patterns may reflect universals of register variation.

In spite of the number of crosslinguistic studies that analyzed discourse, relatively few have examined Brazilian Portuguese-English contrasts on an applied linguistic context. In addition, most existing studies that have addressed BP are unpublished, which makes access to information about BP discourse difficult. The next sub-heading reviews some of those few studies (e.g. Dantas-Whitney and Grabe, 1989; Almeida, 1984; Oliveira, 1997a; Oliveira, 1997b; MORAIGNE e SILVA, 1991; Mc Clearly, 1982).

Portuguese

Following Biber (1988) and Grabe (1987), Dantas-Whitney and Grabe use the Multifeature Multidimensional Framework (Biber, 1984, 1986) to investigate how similar expository texts in BP and English are. Particularly, they compare ten newspaper editorials from the *Christian Science Monitor* and ten from *Jornal do Brasil* on the basis of syntactic and cohesion measures. Among the syntactic variables they analyze are nominalizations, words per sentence, and locative adverbs. The cohesion measure was repetition. All features were hand-counted throughout the twenty-text corpus and raw scores were later standardized for text length. Results indicate a significant difference between English and Portuguese editorials along the abstract/logical versus situational information dimension. English editorials present formal, detached style while Portuguese editorials are significantly more concrete and colloquial. In other words, Brazilian Portuguese shows more frequent use of third person pronouns and locative adverbs than English, and mark concrete content and informal style. English editorials use more

English, and mark concrete content and informal style. English editorials use more nominal content and learned style than the BP corpus. Results related to information orientation were not statistically significant. The authors caution that in large data sets those may reach statistical significance. In other words, high personal involvement and real time constraints may distinguish English and BP editorials. Also, Dantas-Whitney and Grabe demonstrates that English editorials in their corpus use less repetitions, more nominalizations, and ambiguous constructions (e.g. hedges, pro-verb DO, and subordinators) than BP. Results for sentence length were not disclosed.

Though an important first attempt to use Biber's (1988) methodology in crosslinguistic research and advance studies in contrastive rhetoric (Connor, 1996), this study was exploratory and has some pitfalls. A qualitative analysis of the attributes contributing to each dimension seems to be a step that should accompany the quantitative analysis. For instance, qualitative analysis could demonstrate that because Brazilian Portuguese is morphologically rich, it is not as prone to ambiguity in subordinating constructions as English. In addition, the features and dimensions Biber uses derive from large sets of English texts and may not fit Portuguese. Perhaps, as Lux and Grabe (1991) indicate in the case of Spanish, a multivariate analysis on large sets of Portuguese texts would provide a model of factor dimensions for Portuguese that differs from that for English.

This is what Oliveira (1997a) accomplishes. She analyzes 270 compositions (90 in BP, 90 in English, 90 in English as a foreign language) that were written by university students at Pontificia Universidade Católica do Rio de Janeiro (compositions in BP and EFL) and at the Northern Arizona University (English) using Biber's (1984, 1986)

Multidimensional analysis framework. Particularly, she identifies the textual dimensions that characterize the compositions in the three research conditions; examines how the compositions vary across the dimensions identified for the corpus; characterizes the genre of the compositions; and investigates the cultural differences between BP and EFL texts. A Factorial Analysis on the normalized frequencies for 300 words of the linguistic attributes in each dimension indicated that there are five textual dimensions in the corpus—(1) contextual; (2) interactional; (3) organizational; (4) sentential/intersentential; (5) structural. Dimension number one included variables such as number of words, sentences, and independent constructions in the corpus. Dimension number two included cultural, historical, socio-political, and geographic features; dimension number two included interactional expressions such as personal pronouns, manner adverbs, hedges as well as rhetorical questions; dimension number three included observations of problem-solution, cause-effect, generalization patterns; dimension number four included intersentential cohesive ties within and across sentences, and transition words between paragraphs. Last, dimension number five included number of sentences, words, and T-units per text.

A MANOVA on the scores of the variables for each dimension indicates that Brazilian Portuguese, English, and English as a foreign language compositions differ significantly across the dimensions in the case of some variables. Here, I will attain mainly to results relative to Brazilian Portuguese and English as first languages. For example, context dependency variables yielded the most relevant results in terms of variation across compositions written in English and in BP as first languages. Portuguese compositions marked geographic, historical, social, and economic entities significantly more than

English ones. While native speakers of BP would create socio-political scenarios before tackling the topic, native speakers of English would tackle the topic immediately. The researcher interprets these differences as related to cultural differences.

Specifically, compositions in BP granted contextual entities such as the country's political situation highlighted status. The same holds for the Brazilian economic difficulties vis-à-vis those of the world at large. In other words, texts in BP developed from general endophoric references to specific, topical references. On the other hand, compositions in English made little use of such endophoric references. At the time these American university students wrote the compositions, the Gulf War was about to begin. However, they did not bring that in their discussions of the problems facing American families. Rather, they tackled the topic up front, referring exclusively to immediate family problems. There seems to be a prevalence of individual issues over social, politic, and economic issues in compositions written by native speakers of English.

Furthermore, compositions written by native speakers of English were not only lengthier but also had more shorter T-units than those written by native speakers of BP. That is, while BP compositions adhere to an elaborated style, English compositions adhere to a plain style. Oliveira (1997a) proposes that cultural traits relate to the way Brazilians use linguistic variables to construe texts in the contextual dimension. This interpretation is reinforced by the fact that Brazilian learners of English as a Foreign Language use syntactic features with a frequency that is closer to BP than to English. Among other things, she suggests that elaboration derives from the Brazilian polychronic view of time. According to that conception, time is viewed as points that construe a circle. Like time, syntactic constructions progress in circles to construe a topic. Other aspects she relates to

BP more elaborated style are literary traditions, and audience awareness. Students write on the premises of readers who appreciate an elaborated style.

Finally, variables in the interactional-informational dimension also varied significantly across the research conditions (BP and English as a first language; and English as a second language). Personal pronouns (*I* and *you*) yielded the highest frequencies in English as a first language, marking significantly this research condition for interaction. On the other hand, texts in BP yielded negative scores and marked BP texts as less involved with the readers than English texts. Because Brazilians are well known for their interactive style, these results countered cultural expectations. They also countered previous research by Dantas-Whitney and Grabe (1989). These differences may be related to genre differences (editorials versus students' essays) or to cultural influences related to writing instruction in Brazil and the US. While in Brazil the school system emphasizes formal, impersonal writing, in the US the school system emphasizes a greater interaction between writers and readers.

Oliveira (1997a) constitutes a major step toward an understanding of the intricacies of BP discourse. It not only contributes to the identification of functional textual dimensions in Brazilian Portuguese but also provides qualitative explanations of the findings in context, with the descriptive tools of discourse analysis.

Oliveira (1997b) further investigates the differences in syntactic complexity in a corpus of 90 compositions (30 for each research condition) and found that, in English, simple sentences (one finite independent clause) occur more frequently than complex sentences (one main and one or more subordinate clauses). The fact that English uses simple sentences quite often may explain its plain style. Furthermore, results indicate that

it is the distribution of coordinate and subordinate clauses that contributes strongly to the elaboration/plain dichotomy in discourse style. While the percentage of coordinate clauses is higher in English, the percentage of subordinate clauses in complex sentences is higher in BP. In other words, BP compositions show more arguments in dependent than in independent constructions.

The two other studies that addressed Portuguese departed from textual analysis itself, focusing primarily on if and how writing in a second language is similar to writing in the first language. Almeida (1984) investigated cohesion and coherence in native (Brazilian Portuguese) and non-native (English) academic expository discourse of senior university students in Brazil. Though native and non-native discourse do not differ with respect to types and subtypes of cohesive ties in his data, they differ in frequency of use. Successful written expository discourse in BP contained more explicit logical relations (through conjuncts), better organized sequences of events (through lexical repetition, synonyms, and demonstrative reference), longer distance between cohesive items, and fewer exophoric ties. Such differences were found to correlate highly with the general perceived coherence of texts. To explain the differences, Almeida calls on differences between BP and English. In the case of demonstratives, he shows that BP displays a wider inventory than English. While English divides the deitic space into two dimensions--[Spk] close to the speaker and [Neg.SpK] away from the speaker--BP divides it in three levels--[Spk] close to the speaker, [Awa] close to the addressee, and [Neg. Awa] away from both the speaker and the addressee. Therefore, BP is more specific because it carries special forms for intermediary distances. Also, BP is more specific because it is morphologically inflected for gender and number. Furthermore, the statistic

procedures revealed that coherence is not language specific. Rather, it is a global factor bound to the cognitive experience of the writer.

To provide evidence of the cognitive, social and cultural influences on the writing process of L2 learners, Moragne e Silva (1991) case studied the writing process of a native speaker of European Portuguese composing in English, his second language. Findings reveal a complex interaction between the L1 and L2 in the subjects' composing process. That interaction depends on the language that provides the ideational basis for the writing task. In other words, tasks that drew on knowledge that the subject acquired in L1 led him to use L1 in the composing process and to translate concepts and lexical items into English. On the other hand, tasks that drew on knowledge the subject acquired in L2, led him to directly use content and suitable terminology in L2. Furthermore, while he spoke to the researcher almost exclusively in his L1, he took notes, outlined, and transcribed in L2. Most of all, this study indicates that familiarity with a task, whether in L1 or L2, is what has the largest effect on the composing process. Writers complete tasks with greater ease if that task is familiar.

Moraigne e Silva (1991) though, dealt with European Portuguese, a language that differs from Brazilian Portuguese prosodically and structurally (Swan and Smith, 1987). While EP is a stress timed language, BP is syllable timed. Among other differences, EP carries a high number of Arabic words in its inventory, forms the present tense with "estar a" followed by the infinitive with stative and active verbs, rarely uses phrasal verbs and carries a lexicon inventory that is practically not in current use in BP.

The last study that investigated BP discourse is McClearly (1982). It differs from the other reviewed studies in that it takes distance from the applied linguistics context

toward a more cognitive approach to discourse analysis. McClearly (1982) addresses a segment of a Brazilian Portuguese narrative to test Hopper and Thompson's (1980) claim that transitivity is a marker of foregrounding. To do that, she distributed chapter one of *Vidas Secas* (=Parched Lives), a classic of modern Brazilian literature by Graciliano Ramos, to a group of 40 Brazilian teachers and students at the Department of Translation and Interpretation of the Associação Alumni, São Paulo, southeastern Brazil. Their task was to judge if the clauses obeyed the sequentiality or the importance criteria for foreground. Results of the survey indicate that transitivity correlated with judgements of sequentiality, not with judgements of importance. She argues that transitivity is more appropriately viewed as a result of universal perceptual-cognitive functions than of discourse pragmatic functions. It manifests itself in discourse structure via the role of causal events, establishing temporal sequentiality in narratives. This study will further discuss details about perceptual-cognitive functions in the next chapter.

In conclusion, different researchers investigated how linguistic, social, and cultural phenomena relate to the way people use language. The relation between the two variables seems robust given the breadth of existing studies. Major findings that interest this study include the tendency BP writers reveal to use more repetitions and subordination, fewer involvement markers and exophoric ties, and lengthier sentences than English. These findings apparently relate to genre given that Dantas-Whitney and Grabe (1989) and Oliveira (1997a) found opposite results for the interactional-informational dimension. Also, the review indicated that stretches of language between punctuation marks may relate to the amount of information different languages carry in single words and to the "one-new idea constraint." So far, Chafe (1994) has proved that hypothesis for spoken

Seneca, but Damron (1997) contradicted him in her study of spoken Urdu and Pakistani English. In Urdu, units bounded by punctuation are multi-clausal and have more than one idea. Other languages though remain to be examined.

Texts as translatable relations in families of resemblances

Like discourse analysts and contrastive rhetoricians, translators have to understand a text, show its intentions, recognize its frame and socio-cultural trappings (Barbe, 1992) to accomplish their goals. In fact, "discourse analysis is the basis of any real translation" (Delisle, 1988, p.72). Only after understanding the semantics and pragmatics of a discourse can translators perform their tasks. Besides, translation is also a tool for discourse analysis and contrastive rhetoric. There are some pragmatic nuances that only come to light at the moment of translation. Take the case of irony, nonce forms, foreign expressions, or unfinished sentences, interruptions, repairs, false starts in the case of spontaneous conversation (Barbe, 1992). Therefore, translation studies have also contributed to the existing understanding of written discourse, and, in this sub-heading, I address the major trends in translation theory today. Specifically, I focus on its understanding of equivalence and adequacy.

According to Nida (1964), translation is a dynamic interaction among surface and deep structures of the source and target language texts. This interaction should generate a stylistically and semantically equivalent expressions of the source text that includes subtleties of meaning, emotive values of words, and "flavor and feel" of the original message. Though he uses Chomsky's generative grammar descriptive terminology, he inserts culture, context, semantics in the model, misappropriating many of Chomsky's views. House (1977), for his part adds a functional component to Nida's semantic-

experiential components. According to him, source and target language texts should be pragmatically similar.

Contrary to Nida's (1964) dynamic equivalent response theory and House's (1977) functional equivalence theory, Toury's (1991) model of translation sees equivalence and adequacy as multidimensional, complex textual traits. While Nida and House believe the translated text should be source-text oriented or a dynamic response of the readers of the source text in the source culture, guided by semantic correctness, Toury moves the focus to the target linguistic and experiential contexts.

According to Toury (1991), there are different degrees of adequacy and equivalence. Actually, adequacy, acceptability, equivalence are in a continuum. The more a text constitutes a genuine member of the target language literary system, the more adequate, acceptable and equivalent to the source text it will be. The translator's task is to place the translated text among other texts in the target culture. Acceptability and equivalence might then be understood within a neo-Gricean perspective or Sperber and Wilson's (1986; 1995) framework. They represent the expectations of the audiences to whom discourse is addressed and whose members belong to a given discourse community with unique linguistic, socio-cultural norms.

Likewise, Tirkkonen-Condit (1989) and Kempainen (1988) also argue for translation as a textual exercise that should focus on the target audience and not on linguistically accurate, "semantic" translation. The typicality of readers' potential responses and backgrounds come into play. Finally, Gentzler (1993), in a comprehensive review of translation theories, reinforces that tendency of seeing translations less as an empirical fact—a concrete text as defined by the target culture—and more as a complex set of

translatable relations in a given situation. The translated text is viewed as simultaneously drawing on families of resemblances as well as writing itself into other families of resemblances.

Weise (1988) discusses a number of variants of British and American scientific texts translated into German. His analysis shows that factors governing communicative competence should guide optimum translation of scientific texts. For example, in one of the texts he analyzed, doctors who are members of the medical science discourse community recommended that a sentence be deleted for it was too colloquial for German scientific texts. In English, however, the sentence fit well. Semantic invariance (correctness) combines with pragmatic adequacy (appropriateness) to yield communicative equivalence. Weise shows that socially determined factors such as norms and conventions restrict linguistic choices, replicating findings of Heath (1983), Scribner and Cole (1981) and others cited in my review of contrastive rhetoric studies and ESL writing.

In addition, pragmatic devices such as foregrounding of new information and backgrounding of old information, devices to maintain continuity of the topic, devices for giving prominence to whatever the writer views as most important should guide choices in the TL, narrowing the number of variants. Thus, for Weise (1988) equivalence is a measure of the extent to which a TL text is communicatively congruent with the SL text in content and intent, reflecting the TL typical way with words.

Equivalence has been widely discussed in studies that compared versions of a given text in different languages. Abu-Ssaydeh (1993) examines the difficulty in finding short equivalents when coining new Arabic terms for English words and concepts; Alvarez (1993) addresses the translation of metaphor in Spanish and English; Menacere (1992)

focuses on grammatical and morphological constraints in Arabic and suggests how to render them in English; Doherty (1992) investigates the relativity of sentence boundary between English and German, stating that translators often make the syntactic adjustments required by the language when moving from source language to target language. However, they hardly alter sentence boundaries. This pattern may not always be acceptable as we will see in Portuguese-English translation. The results of the above studies demonstrate that the translator's concern should not be with similarity or dissimilarity of features (Weise, 1988), but with adequacy and appropriateness in the TL version. It is both the semantic and pragmatic contents of the SL text that are reconstructed in the target language. The issue in fact, in Weise's view, is communicative equivalence or the selection of those linguistic features in the target language that are known as typical by the audience.

If we think of van Dijk and Kintsch's text bases (=sets of predications; 1978; 1983) we may see discourse translation in different languages as different arrangements of the same underlying text atoms. The texts would be different textualizations of the same input predications. They arise out of the same predications but are realized through different text strategies and processes of textualization.

If we ask where predications come from, we have to dig for answers in cognitive science. Catford (1969), Bröeck (1978), Wills (1982), and Snell-Hornby (1988) did that, accounting for texts as a communicatively oriented configuration with a thematic and functional-pragmatic dimension. There is need to look at reality, at real texts in the target language and begin the analysis and the translating task from there. Everything can be expressed in every language. The task is to strive toward optimal, typical solutions.

Summary

Accounts of written discourse have exhibited a multifaceted picture. Frequencies of linguistic categories have been counted and related to textual functions; cohesion and coherence have been intensively studied; linguistic structure has been related to cultures, discourse communities and situational contexts, concrete reality, and to concepts that users of different languages share.

The review in this chapter shows first that most studies that have examined the frequency of linguistic categories and their functions, psychological notions such as topic continuity, activation cost, and processing effort have dealt almost exclusively with English only. Chafe (1994), Biber (1995), Dantas-Whitney and Grabe (1989), Oliveira (1997a), and Oliveira (1997b) are exceptions.

Second, it shows that writing reflects a selection, by a writer, among different linguistic devices. That selection is guided by writers' intentions, writers' assumptions about readers knowledge and relevant information, typical properties of discourse type and genre, conventionalized social and cultural writing norms, and processing effort.

Third, it shows that there is a pervasive link of grammar with discourse communication. Linguistic devices function as cognitive instructions that accelerate or retard comprehension.

Fourth, it defined written language as a socio-cultural, cognitive phenomenon that, as much as spoken language, is situated, interactional, and constrained by the limited capacity of working memory..

Fifth, the review established the importance of cognitive structures to comprehension. Texts are accepted as successful when they fit into the addressors'

schemas. Notions such as discourse appropriateness and adequacy depend on how well the arrangement of textual atoms conform to predications that are conceptual and that are shared by the target speech community. They also depend on how those predications are actually realized in pragmatic terms.

Despite the great number of studies (this dissertation covered only those directly related to its immediate scope) and the relevance of cognitive structures in comprehension, there are no analyses of texts that combine advances in discourse analysis to recent theoretical advances by cognitive linguistics and Cognitive Grammar. For instance, one such study could analyze grammatical categories as metaphors or motivated conceptual structures. To enrich the body of research the cited studies typify, this dissertation combines the descriptive tools advanced by discourse analysis studies with those advanced by cognitive linguistics, specially by Cognitive Grammar as conceived by Langacker (1987, 1991). It will use those descriptive tools to discuss central aspects of a text image-schema in English and Brazilian Portuguese.

In addition, because this study investigates a language other than English, it will, deepen our understanding of the differences between written and spoken discourse across languages. Features that have been typically associated with spoken discourse (e.g. fragmentation) in English, may be typical of written discourse in other languages like Portuguese that are morphologically rich. The same is true of genre analysis. Some features of specific genres may hold in English, but they may vary crosslinguistically.

Another important contribution this dissertation may offer relates to Chafe's (1994) and Damron's (1997) findings about typical intonation units. Chafe's findings about spoken Seneca confirm that intonation units in morphologically rich languages tend to be

shorter than in English and generally conform to the “one-new idea constraint.” Damron’s findings about spoken Urdu negate Chafe’s hypothesis. Though this dissertation deals with written language, it may throw some light on how written Portuguese compares to English in terms of punctuation units and the adherence to the “one-new idea constraint.” If you consider the role of grammar in guiding information processing, it may be even the case that written punctuation units in BP will be lengthier. This study will contribute insights in that direction.

The view of grammatical categories, mostly morphological trappings, as mechanisms for discourse comprehension is the basis of the analysis I will conduct. With the appearance of cognitive linguistics in the 80’s, textual analysis that follows that path are in order. The next chapter discusses the cognitive approach to linguistic description, presents the main principles that guide it, reviews studies that adopted it, and shows how cognition governs discourse.

CHAPTER III
THE COGNITIVE APPROACH

Overview

A cognitive approach to linguistics highlights the role of experience and perceptions as sources of knowledge, and refuses the notion of linguistics as an autonomous faculty. In addition, it applies no clear boundaries among syntax, semantics, pragmatics, and morphology. Those systems work together to construe meaning. Linguistic categories, thus, are instruments for conceptualizing and interacting with the world. Another characteristic of the cognitive approach is its concern with constructions as the fundamental unit of syntactic description. Among others, there are cognitive accounts of the *there* construction and the preposition *over* (Lakoff, 1990), *possessive* constructions (Taylor, 1989; Langacker, 1991), nouns, verbs, and the passive in English (Langacker). All these accounts explain grammatical constructions as overt manifestations of experiential gestalts in the light of pragmatics, semantics, syntax, and morphology. Finally, the cognitive approach sees language phenomena as products of various cognitive mechanisms. Among them, Idealized Cognitive Models (ICM's), image schemas, mental spaces, radial categories, and metaphor mapping. According to the cognitive approach, two forces collude to group elements in linguistics—shared features and radially structured categories. This dissertation will deal with shared features.

The studies treated in this chapter illustrate and further develop those basic tenets. They are divided into five sections that reflect the fundamentals of cognitive linguistics. The first section, principles of categorization, explores how mental, perceptual and physical attributes conspire to form “the categories we live by” (Lakoff, 1990); the second describes one of the cognitive mechanisms that produces language phenomena—image-schemas—addressing papers by Watters (1996), Delbecque (1996), and Borneto (1996). The studies in the third section discuss grammatical constructions invoking the notions of family resemblances, motivated categories, image-schemas, and metaphorical extensions. The fourth section, aspects of Cognitive Grammar, describes the notational tools developed by Langacker (1987; 1991) and applies them to the analysis of nouns, verbs, word order, nominalizations, locative expressions, grammatical markings, and modifier-noun sequences. The last section addresses the relation between language, thought, and culture or the Sapir-Whorf Hypothesis, underscoring a psychological interpretation of the weak version of the Hypothesis.

Principles of categorization

Category formation derives from functional and psychological aspects (Rosch and Mervis, 1975; Rosch, 1978). Functionally, categories are classes of related phenomena grouped together mentally on the basis of some kind of salient perceptual or functional similarity (see Györy, 1996, for a historical perspective of categorization). For example, chairs, tables, sofas are exemplars of the category furniture. The similarity among the members of that category arise from individual judgment calls, not from objective characteristics of the entities themselves (Taylor, 1989; Lakoff, 1990). It is the way individuals interact with an environment that determines if they categorize a tomato as a

vegetable or a fruit. Labov (1973), in a study about house receptacles in which subjects were asked to name if objects were cups, mugs, bowls or vases found that shape, size, material as well as the use to which the objects are put influence how receptacles are labeled. Casad (1988), while investigating Cora, a Uto-Aztecan language of Northwest Mexico that is permeated with topographic adverbs, locative particles, verbal prefixes of location and direction, correlated its structure, to the geographic and social environments. In Langacker's words (1991), "the meaning of an expression is not given solely (if at all) by the objective properties of the situation it describes—rather it is a function of how speakers construe the situation and structure it by means of specific images" (p.56).

Psychologically, categories provide maximum information with the least cognitive effort, structuring incoming information from the perceived world according to pre-existing experiences (see also Neisser, 1976; Langacker, 1991). The new structures are easily learned because they are metaphors of the old ones. It is our knowledge of the source domain (the old experiences) that motivates our understanding of the target domain or of the metaphor expression (the new information). In Cora, the *u/a* (=inside/outside) central asymmetry motivates multiple related meanings. It may indicate location on an inner vs. outer surface as well as mark deep vs. shallow penetration to the interior of the surface. It also accommodates the contrast inaccessible/accessible which relates to inside/outside because an entity is always out of view when contained inside another (Langaker, 1991).

Specifically, the formation of categories obey five principles (Rosch and Mervis, 1975; Rosch, 1978): (1) If members share few attributes of low cue validity, they form a superordinate category such as furniture, vehicle, fruit, and clothing; (2) The more a

member has attributes in common with other members, the more likely it would be considered a prototype or the best exemplar of the category. According to my experiential gestalt, the most prototypical member of the category vehicle is car and the least elevator; (3) The more prototypical a member of a superordinate category, the less likely it is to dominate in categories other than the one in question; (4) Basic categories share many attributes, being inclusive, abstract, and rich in that there are attributes common to all or most members of the category; and (5) More prototypical members are easier to learn and identify.

Different studies corroborate the fifth principle by demonstrating that cognition interacts with category formation. Armstrong et al. (1983) found that folk subjects and expert mathematicians do not assign the same status to numbers. While folk subjects considered the number *three* a better example of an odd number than 801, expert mathematicians did not. Their formal education imposed definitions and criteria on their understanding of basic numbers that extended their image-schema of the category. Brown (1973) shows how children acquire prototypical members of a category earlier than other members because of perceptual gestalts. For example, past tense forms that refer to immediate past (central sense) emerge before counterfactual past given their noticeability, meaningfulness, and distinctiveness. The same is true of motion verbs (Miller and Johnson-Laird, 1977). Because they are conceptually dominant, grounded in daily experience, and the most frequently used ones, they are learned early and serve as the base for several metaphors such as CHANGE IS MOTION (Radden, 1996).

Györy (1996), in his analysis of semantic change and the process of category formation, demonstrates that category formation is a historical and cultural process that

may extend over generations. Diachronically, there is a natural evolution or redefinition of attributes in a category because category formation depends on the way the world happens to be and what we happen to know about it. The incorporation of new members occurs by extension from prototypes. The main reason for incorporating new members is successful communication. As the discussion in chapter two shows, understanding only takes place when the same cognitive structures that are activated in the addresser gets activated in the addressee (Lakoff, 1990; Györi). Borrowing Györi's words, it is through a common model of conceived reality that we share individual concepts of experiences that will function as recognition devices for instances of a category (p.177).

MacLaury (1995) combines functional and psychological aspects of categorization. He emphasizes that a category is a vantage, a point of view that people establish and alter depending on external motivation to comprehend the world. That model of categorization, the Vantage Theory (1995), highlights the role of the categorizer in arranging the coordinates of an image or reference point and in selecting those coordinates that facilitate comprehension. According to it, we form categories to establish a view of the world; to reduce variation to manageable groupings; and to create and refine our viewpoints.

In conclusion, human categorization is a context-dependent phenomenon (Barsalou, 1992) that depends on how the conceptualizer construes the elements available in working memory. People categorize to ease the processing effort. Though there is no single collection of properties that all members of a category (or family) share, they resemble one another in different ways—physically (shape, material, size), functionally, and perceptually. In MacLaury's (1995) words, "everyone sees the same world, but different

people emphasize or suppress their awareness of different parts of it" (269). When people consistently emphasize certain attributes with high frequency, they form a fixed gestalt that will function as a reference point or a schematic image.

Understanding image-schemas

Investigations about discourse comprehension often place scripts, schemata, frames on the spotlight (see Chapter II) because of their facilitative roles. Cognitive linguistics research that deals with natural language phenomena is not different. However, within cognitive linguistics the traditional schemata (Rumelhart, 1980) are more similar to ICMs or general cognitive structures that organize our knowledge. For example, the ICM OF ORDINARY COMMUNICATION that structures social and linguistic interactions (Sweetser, 1984). ICMs differ from schema theory (Rumelhart, 1980) in that ICM's can account for the interaction of different cognitive models, including one that is personal and exclusive of the conceptualizer. Therefore, while ICMs can account for variations within cognitive models by means of these interactions, schema theory cannot. Those very interactions are the phenomena that give rise to prototype effects within categories. An additional difference between ICMs and schemata, as defined in reading theory, is that ICMs have different organizing principles. One of them is an image-schema, a cognitive semantics term, coined by Lakoff (1990), to characterize structures that arise from physical and social experience. According to this understanding, image-schemas are more concretely embodied than traditional schemata.

Image-schemas may be defined as abstract conceptual structures (such as the part-whole, cause-effect, center-periphery, container, source-path-goal) that model perceptual invariants and have a rich internal logic (Johnson, 1987; Lakoff, 1990; Langacker, 1995).

For example, the MOTION image-schema is notationally represented by the structures SOURCE, PATH, GOAL and DIRECTION (Radden, 1996). Those structures are motivated by our bodily experiences. "Every time we move anywhere there is a place we start from, a place we wind up at, a sequence of contiguous locations connecting the starting and ending points, and a direction" (Lakoff, p. 275).

In addition, image-schemas serve as concrete reference points for motivated metaphorical mapping onto abstract domains. For instance, the CONTAINER schema (a boundary distinguishing an interior from an exterior) extends our bodily understanding of *in* and *out* to abstract domains such as in "walking out of a deep sleep" or "reaching into the medicine cabinet" (Lindner, 1981). The PART-WHOLE schema is present in our understanding of institutions, families, society (Lakoff, 1990). Lakoff emphasizes that schemas are not meaning postulates, but *gestalts*, structured wholes that are more than the collection of their parts and that structure our direct experiences. Even if the collection of parts exist, they may not constitute a whole. A family is a good example. The father, the mother, and the children constitute a family if they are together as a whole. In the case of divorce, that whole is split up.

According to Lakoff (1990) and Deane (1996), grammatical structure is a metaphoric projection of these schemas, being equivalent to physical objects. The entire grammatical system is based on the processing of fundamental grammar relationships as if they were linkages between physical objects whereas linkages mean functional dependency, co-occurrence. It follows that accounts of image-schemas normally include factors such as vantage point on a scene, involvement, and the orientation of foregrounded entities in relation to backgrounded ones in a given context (Casad, 1996).

Typical cases are the tentative image-schemas proposed by Delbecque (1996) for the prepositions *por* and *para* (=to) in Spanish and by Borneto (1996) for the locative verbs *stehen* (=to stand) and *liegen* (=to lie) in German. *Por* and *para* schematic images correspond respectively to the notions of “CONDITIONING” and “PROJECTION,” according to their distribution in a corpus of essays and to the range of related meanings they take in different scenes. The superordinate node holds the abstract schematic meaning. From that node, other conventionally sanctioned uses originate by chaining, as a result of the different profiles each of them impose on the base as well as of how language users construe the scene. While *por* marks the prepositional object for pre-conceptualization vis-à-vis the prepositional subject, *para* does not. Rather, in the case of *para* the conception of the prepositional object is a “projection” of the prepositional subject. The evidence for that conclusion originates in the distributional differences of *por* and *para*.

The locative predicates *stehen* and *liegen* in German (Borneto, 1996) correspond to the image schemas BEING LOCATED VERTICALLY (*stehen*) and HORIZONTALLY (*liegen*) as they appear in a corpus of newspapers of the Institut für Deutsche Sprache-Manheim. The conceptual content of the VERTICALLY LOCATED schema for the verb *stehen* evokes images such as HUMAN VERTICALITY, GEOMETRIC VERTICALITY, GRAVITY VERTICALITY, and SALIENCY VERTICALITY that not only share common characteristics but that depend on each other. For instance, every instantiation of VERTICALITY or HORIZONTALITY involves some kind of movement and one leads to the other. Gravity verticality evokes geometric verticality and so on. Borneto provides a general explanation of all the usages of the verbs, basic and extended,

in cognitive terms. Particularly, he highlights the links between the basic usages and the respective image-schemas. Like in Delbecque (1996), those links relate to the perspective and involvement of language users in a scene.

Another interesting aspect of Borneto's (1996) research is his investigation of the prototypical instantiation of the image-schema. He designed a test in which he listed 40 objects based on shape and orientation and asked 15 native speakers to match them with *stehen* or *liegen*. He also asked native speakers to indicate objects that would match with both verbs. Answers proved consistent with his analysis and underlined the centrality of the human body and life experiences to semantics. Both the frequency (perceptual experience) and the social relevance of the entities played an important role in the native speakers' choices.

In brief, to get at an image-schema means getting at the thematic ground, the cognitive anchor of the construction in focus. The image-schema of a text consists then of a description of a set of attributes, their interrelations and thematic ground where each attribute can be instantiated by information in the text or from default information that is part of the image-schema (Smith and Swinney, 1992, p. 303). Extensions from it derive more or less marked members in relation to a prototype.

Linguistic devices, discourse, and cognitive linguistics

In cognitive linguistics, grammatical phenomena are much like any other category. being subject to the same principles of categorization. Like in a network or in family resemblance relations, the structure of grammatical phenomena presents prototypical effects, with more central and more marginal members. That is, nouns, prepositions, words, articles, transitive predicates, noun phrases, subjunctives, possessives are

categories which have no clear cut boundaries but are sensitive to functional and psychological gestalts, being explained in terms of image-schemas.

Linguistic devices. Langacker (1991), for example, defines a noun as “a linguistic unit which profiles a thing where thing is defined as a region in some domain” (p. 17). By placing grammatical phenomena in the sphere of concepts, he analyzes nouns as entities which bear some type of meaning relation in a domain. Consequently, the different meanings that a NP may assume within the noun domain can be hierarchically ordered. We may talk about most “noun-phrasal” NPs (Taylor, 1987) like *the boy* or those which refer to conscious volitionally acting, animate creatures and about least “noun-phrasal” NPs which range from concrete inanimate creatures to the deictic *there*. The same is true about affixes and words. *THE*, as discussed by Taylor (1989), is a “clitic-like word” because it has more word-like attributes than affix attributes, including freedom to attach to anything.

Also, as mentioned before in this chapter, category extensions occur mainly in terms of metaphors as defined by Lakoff (1990). It is by projecting the thing-schema onto non-spatial domains that we account for the nounhood of *departure*, *color*, and *pitch*. Likewise, it is by projecting the temporal domain on to the thing-schema that verbs are de-categorized. Taylor (1989) says that they lose the morphological trappings of their class (see for example *eat* in *Natália is ready to eat* where *eat* cannot realize any of the contrasts typically associated with verbs), becoming nominals.

Further exemplifying how cognitive linguistics pairs grammatical constructions with prototype effects and radially structured categories, I now discuss and evaluate the

treatment of four grammatical constructions: possessives, subjunctives, the word LIE, and the morphological class *string/strung*.

According to Taylor (1989), possessives semantically identify the “possessed” with reference to its “possessor” as in the phrase *Tiago’s skate*. The possessor, as indicated by Langacker (1991), must be construable as a “reference point entity” or a landmark (LM), a path we need to trace to locate the possessive or “thing.” This is the center sense of the possessive construction. However, it is a cluster of attributes that grants admission to the possessive category. For example, the possessor has to be preferably a human being such as *Tiago* or an animate. Inanimates like in *Today’s humor* are less typical possessors because *today* cannot possess things. Also, the possessed entity tends to be a concrete thing. That is another attribute the example above fails to present (*humor* is not a concrete entity). Taylor lists other attributes of the possessive category, including that the possessor has rights over and duties with the possessed and to exercise them the possessor needs to be close in spatial proximity to the possessor.

However, as I mentioned before, the cognitive paradigm gives room for extensions from the prototype. Some instantiations will be further and others closer to the cluster of attributes of a prototypical possession relationship. For example, in *the table’s legs*, we do not have an animate possessor, but an expression that relates a part to a whole, perspectivizing the proximity between the possessor and the possessed, the exclusiveness and long-term relation between a thing and its properties. This is then a marginal member of the category as much as *the car’s engine* and *the dog’s ears*. Climbing up the scale of membership, we also have possessives like *the analysts’ computers* in which the relation between the entities diverge from the prototypical because *the analysts* may not actually

have rights over *the computers* though they have duties and may use them temporarily. Besides, the relation between *the analysts* and *the computers* is not a long-term one, but a limited and non-exclusive relation. Another instance, now even closer to the prototype is *the dog's house*. Though *dog* is not a prototypical possessor (it is not a human being), it takes all other attributes of the category *possessor*. Furthermore, the relation between *dog* and *house* reflects the exclusive nature of the relation between *possessed* and *possessor*. That leads us to another use of the possessive: to identify particular entities. *Jacques' office*, for instance, refers to a unique image-schema. The same is true of kinship and other personal relationships such as *Tania's spouse* or *Jerry's associate*. Similar interpretations may account for possessive constructions like *the company's director*, *the state's governor*, or even more interestingly, *the plane's departure*, *yesterday's bombing*, and the like. In those possessive constructions the abstract entities are specifically located with respect to another entity, a participant or circumstance, and they constitute examples of marginal members of the possessive category. Some other more marginal members may give place to idioms like *the year's end* or less acceptable expressions like *the moon's shape*. These marginal expressions do not permit rewordings as in **This shape is the moon's* or **this bombing is yesterday's* which are totally possible in more prototypical examples such as *This game is Rafael's*. Further extensions would prove unacceptable like in **the arrival's time*.

In brief, there are numerous instantiations of the possessive category and they stand in family resemblance relations. Basically, the generalization we can make about that category is that it profiles a relation between two entities in which the possessor has some primacy over the possessed. Also, as Taylor (1989) points out, the possessor should not

diverge too much from the prototype--a human being--otherwise full productivity of the category is replaced by idiomaticity, doubtful acceptability or even unacceptability (p.193).

Subjunctive constructions also stand in family resemblance relations as discussed by Winter (1989) when investigating the subjunctive mood in French. In old French, the subjunctive prototypically expressed indefiniteness. This image-schema manifested itself as negativity, subjectivity, and most centrally doubt. That is the case of sentences like: *Ge ne cuit pas que gel connoisse* (=I do not believe that I know him; Winter's example) in which overt negation is used to trigger the subjunctive. There were also cases in which covert negation involving comparisons of inequality or some kind of concession equally triggered the subjunctive, as much as expression of fear which would instantiate uncertainty. The fact of the matter is that instantiations of the subjunctive could be explained under the umbrella of uncertainty of outcome or lack of definiteness as in any other radially structured category.

Though some of the uses of the subjunctive in Modern French correspond to those of Old French, the prototype shifted and instead of doubt, French speakers now place subjectiveness or judgment at the center of the category, having admitted new members to the category. For example, French speakers now use subjunctive constructions after expressions of emotion and assertions (*c'est un fait que* -'it is a fact that'). Also, while wishes and commands still take the subjunctive, hypotheticals no longer do, and restrictives, expressions of future intention or goal, and expressions following triggers like *jusqu'à ce que*, *après que* have been admitted to the category. In questions of opinion, by extension, the subjunctive is also spreading (subjective sub-category) as much as in

expressions of emotion or appreciation like *Je suis contente que tu sois ici* (=I am happy that you are here).

In brief, the subjunctive category in French has extended its domain based on the cue validity of some types of modal choice and on the development of category extensions. It has also lost some marginal members to the now central notion of subjectivity or personal opinion. The new members (uses of the subjunctive after assertions or emotion expressions) are extensions of the new prototype and because of their meaningfulness, they emerge as signals of productivity.

However, Winter (1989) and Taylor (1989) failed to pair pragmatic constraints governing the use of the subjunctive and the possessive constructions with syntactic and semantic notions. Coleman and Kay (1981) have contributed in that direction while investigating the word LIE. They hypothesized, among other things, that an utterance containing just one of the characteristics of a prototype *lie*--falsehood, deliberateness, and intention to deceive on the part of the speaker--would be considered less a lie than an utterance containing those attributes plus one other. It turned out that the more prototypical "lie-like" attributes a story contained, the higher it scored on the lie scale. In addition, for each pair of stories that differed in number of prototypical attributes, the one containing at least one more element than the other scored highest on the lie-scale. As the authors say, the use of words like *lie* depend on prototype effects but also on knowledge of the occasions, reasons, appropriateness for classifying an utterance as such. They correctly suggest that classifying an utterance as a lie is part of the ethnography of speech communities. That is, the responses subjects provided reflect their cognitive assessment of

how well the story fits the prototypical lie (a cognitive image), as well as their ability to judge it as such which is motivated by other world knowledge or pragmatic related facts.

Bybee and Moder (1983) reinforced that linguistic categories show prototypical effects while studying the class of strong verbs *string/strung*. They hypothesized that subjects in their sample would produce more past tense forms with /ʌ/, the closer the constructed nonce forms the authors provided were to the prototype. Using historical and experimental evidence, Bybee and Moder also hypothesized that the class has a product-oriented prototype with the following characteristics: it would begin with an /s/ C (C)--; end with a final velar nasal /ŋ/ or /ŋk/ (since there are more verbs with /ŋ/ they surmised that it would make a verb closer to the prototype than /ŋk/); and have a lax high front vowel /ɪ/. They theorized that based on lexical associations among existing past forms (“schemas” or generalizations about the shape of a past tense based on the existing lexicon), speakers would produce past tense forms by chaining. More prototypical members would then share a greater number of the listed characteristics (e.g. *string*) than marginal members (e.g. *drag*). Results confirm their hypotheses. All phonological parameters affected the likelihood of membership in this verb class and boundaries among members are in a continuum.

This study argues against a theory of language competence that separates cognition from linguistic competence and performance. Prototype effects, borrowing Taylor’s words (1989), permeate the very structure of language, originating, among other things, from the way concepts are structured in the mind and put to use in everyday life. Schemas that associate words based on phonological shape and morphological function simultaneously structure the class *string/strung*, according to this study.

Other studies such as Shen (1996) and Floyd (1996) have dealt with grammatical categories establishing a network of related senses within a prototype structure. Shen (1996) examines the main usages of the verb *lai* (=to come) in Chinese and its complements and draws two conclusions. First, that different meanings depend on sentence type. Second, that different semantic structures are interrelated by a shift from the domain of physical space to that of mental space or from an objective to a subjective perspective. The different semantic structures of *lai* are grouped into those that specify spatial motion and those that specify abstract motion.

Floyd (1996) investigated the reportative suffix *-shi* in Wanka, a dialect of Peruvian Quechua. He shows that, prototypically, *-shi* suggests that an utterance is based on hearsay. This central prototypical usage as well as one or more of its extensions motivate different instantiations of *-shi* and form a radially structured category. Other usages include marking the source of authority for folklore, in riddles, or as a challenge structure. Though there is no central attribute that all usages of *-shi* share, together they constitute a radial category, involving inter-linked variations of the notion of information that arises from the involvement of other participants.

Thus, analysis of grammatical phenomena in the light of cognitive linguistics allows category boundaries to be flexible, constructions to be metaphors, and networks of related constructions to cluster into an image-schema or to be prototypically structured. In doing so, cognitive linguistics resolves conundrums we could not satisfyingly explain with classical approaches, such as calling *management* a noun or constructions that have no typical agents, and provides new avenues to the analysis of discourse. In the next sub-

heading, this dissertation reviews studies that have applied the notions developed by cognitive linguistics to the analysis of discourse.

Discourse. In his introduction to the volume *Cognitive Linguistics in the Redwoods*, Casad (1996) mentions that “the life of a redwood forest is found within the morphology of its architecture” (p.6). Then, he draws on *Cognitive Grammar* (Langacker, 1987; 1991) to explain that all language units fall in a continuum of schematic structures of different lengths, from morphemes to lexical items to phrasal structures, onto sentences, and up to discourse. As we move upwards in the complexity scale, the more schematic and less conventionalized the patterns tend to be.

Discourse then, is one of these highly schematic structures that within cognitive linguistics could be interpreted as the Redwood Forest, borrowing Casad’s (1996) metaphor. To analyze it, we should look into its life, that is, morphology and syntax, within the perspective of cognitive grammar (Langacker, 1991, see the next heading).

Though studies within the cognitive linguistics paradigm analyze different structures in the context of discourse, they neglect the analysis of discourse as a “Redwood Forest.” In my review of the literature of existing accounts of written discourse, Rubba (1996), Duszak (1995), Snell-Hornby (1988), Neubert (1985), and van Dijk (1988) ventured into that domain, talking about texts as prototypes or prototype communicative behaviors. According to their views, texts are holistic experiences that cluster to form a category. That understanding proves invaluable for crosslinguistic research. As we have seen in the review of the studies that have addressed texts as linguistic, social, and cultural phenomena, and in my review of translation studies, different cultures have different expectations about content and the rhetoric structure of texts. If we conceptualize texts as

a prototype structured category, links among cultures will naturally emerge as well as differences.

Duszak's (1995) crosslinguistic investigation of prototype effects in news stories is particularly illustrative of that point and relevant to this dissertation. Addressing English, German, and Polish news stories, she presents a news story prototype for Polish and German and contrasts it to van Dijk's (1988) news story prototype for English. Her central claim is that there are different expectations about what is natural, effective, and aesthetic in news stories because of different prototypes.

To prove her claim, she collected 87 German and Polish reports on political and social events that appeared in newspapers comparable to *The Washington Post* and *The New York Times*. Then, she examined them for level of segmentation, paragraph structure and the flow of narration (dominant thematic lines). Results indicate that Polish and German texts deviate from van Dijk's (1988) English news-schema prototype. According to van Dijk, the basic mental schema for English news stories is "somebody-did-something-to somebody for some reason, somewhere, some time." An episodic model triggered by the headline and elaborated by discontinuous themes grounded in pragmatic coherence guides the comprehension of that schema.

Whereas the prototypical American English newswriting is dominated by telegraphic pattern of content selection and organization, German and Polish newswriting centers narrative techniques typical of regular storytelling and at some techniques of expository writing. Compared to English, Polish and German texts were less segmented, showed more integrated paragraphs (longer than English ones) with explicit cohesive links, initial build-up of the setting, chronological cueing, longer sentences, elements of

historical and contextual background, evaluative or predictive comments, and complex sentences with embeddings and elaborations.

Duszak (1995) grounds the differences on the structural format of the languages, historic literary traditions, and the public communication style valued by the speech communities in question. Among other facts, because of inflections, sentences in Polish or German are highly expandable, freely allowing right branching constructions and layers of nominalizations. The gradual development of the message coincides with an increase in communicative importance. In addition, German and Polish news texts are marked by European traditions of literacy and crosslinguistic cross-cultural contacts. Borrowing Duszak's words, "they resemble reports made by a perceptive onlooker rather than news dispatches" (p. 477).

An additional contribution to the DISCOURSE IS A RED WOOD FOREST metaphor, comes from Rubba's (1996) analysis of case marking for verbs of cognition in German texts. Contrary to Duszak (1995) and van Dijk (1988), this pilot study uses the descriptive tools of Cognitive Grammar (CG), as proposed by Langacker (1991; see the next sub-heading in which the fundamentals and terminology of Cognitive Grammar are explained). Rubba's study relates syntactic devices in the texts to conceptualization and construes discourse as an image-schema or a folk model (complex cognitive schemas, imprinted in culture, from which people draw to understand the world around them and to communicate). Her data come from two interviews in which she asked two native speakers of German for differences between pairs of mental experience verbs such as *to think* and *to doubt*. One member of the pair would take an accusative object and the other a dative object. It could also be that one or both members of the pair would occur with a

preposition. The basic question the study addresses is why some verbs of mental experience govern only accusative if they are followed by prepositions that can take either the dative or the accusative.

Rubba's (1996) findings indicate that the metaphorical structuring of experience in a folk model motivates the choice of structure. That is, the conceptualization of the event that the verb encodes and that experience has engraved in the minds of speakers determines the selection of case. Though her results derive from interviews with only two informants and require further investigation, case marking appears to be directly entrenched in cultural models and in the conceptualization of language.

Similarly, Verhagen (1996) relates the order of elements in a sentence to sequential conceptualization in his investigation of epistemic verbs like *promise* and *threaten* and relative clauses in the Dutch Eindhoven Corpus. He has examined specifically why epistemic verbs in Dutch occur to the right of and adjacent to the complement (c.f. * *It promises tomorrow a fine day to become*; Verhagen's example, p.798). In addition, he has investigated why relative clauses are extraposed (c.f. * *A subject must be chosen that will defer the drudgery of actual writing till death*; Verhagen's example, p. 798) or separated from their antecedents by a verb (SOV order in German and Dutch).

Verhagen (1996) claims that word order results from a subjective characterization of the referent and the argumentational orientation of discourse. In other words, vantage point (see also MacLaury, 1995, on Vantage Theory). If the verb comes first, it is because it needs to be conceptualized independently and construed objectively by the conceptualizer; if it does not, it is because it indicates the lack of relevance of the process and the relevance of orientation. In the first example, *promise* has to be conceptualized

with the predicate nominal since it does not provide any information about the subject independently of the rest of the sentence. On the contrary, the sequence *promises to be a fine day*, as Verhagen says, provides a subjective characterization of the referent of the subject. The same is true of the extraposed relative clause. The referent of the object NP and the process denoted by the verb are to be conceptually integrated. In Verhagen's words (p. 793-794) "whenever the elements in a sentence are distinguished as separate, the one that comes first is to be conceptualized independently with respect to the one that follows."

In brief, all of our linguistic knowledge is organized and stored in cognitive categories (Janda, 1996) or a network with clusters of patterns that bear markedness relations. Texts are no exception. Unmarked elements of the image-schema "text" will give the category its shape, determining what is the most prototypical, expected, and inclusive element that should occupy the central position of the category. Marked elements will fall in a continuum from most to least marked in relation to that prototype. A close examination of the patterns (morphology, syntax, orientation, information structure) may reveal language specific image-schemas for written discourse. This is the purpose of this dissertation.

Summary. The cognitive approach accounts for linguistic devices and discourse in terms of prototypes, image-schemas, clusters, claiming that grammatical phenomena group according to fundamental semantic properties (see Langacker, 1991; Lakoff, 1990; and Taylor, 1989). Because these properties are embodied in perceptual experience and functional life, they ease the effort of processing and are facilitative of comprehension. Form derives from aspects of meaning and pragmatics. It follows that in a category or

domain, the more peripheral members tend to be instantiated more sporadically and unpredictably than the more central members. As such, the more peripheral members are harder to process. In contrast, the more central members of a category are more frequently used, therefore more salient and less costly to comprehend.

Relevant Aspects of Cognitive Grammar

Because this dissertation analyzes Brazilian Portuguese and English discourses with fundamentals and descriptive tools typical of both discourse analysis and Cognitive Grammar (Langacker 1991), I briefly describe here the aspects of CG that are most relevant to the work I present later. In addition, this sub-heading presents nominalizations, word order, head-modifier sequences, and other linguistic expressions within the perspective of CG to ground notions and vocabulary that will be used in the qualitative analyses (I also direct the reader to the glossary in Appendix A).

According to CG, form, meaning and conceptualization are one and the same and should be accounted for in terms of cognitive processing, sensory, emotive, and kinesthetic sensations. Linguistic expressions are then a function of cognitive domains or mental experiences such as folk models or irreducible domains like space, time, color, and motion or even highly abstract image-schemas like those previously described in this chapter. To attain semantic value, linguistic expressions impose a *profile* on a *base*, that is, they make prominent a sub-region in the knowledge structure that is active and that relates to the expression in a given context. For example, *eyes* profile a region on a base that is the *human body*. The profile-base distinction also relates to a notion that, though less complex than image-schemas, is equally important linguistically and cognitively (Langacker, 1995)—conceptual archetypes. In Langacker's view, those archetypes

constitute our conceptions of the human face, body, objects, agent-patient interactions, face-to-face conversations, affecting other entities with an instrument, physical motion through space, and so on. While image schemas characterize schematically all instances of a category, a conceptual archetype characterizes the prototype.

For instance, Langacker (1991) characterizes a *noun* as a thing whose archetypal conception of a physical object provides the category prototype; a *subject* as our ability to impose figure/ground organization on a conceived situation; and a *verb* as an event construed as a temporal process or relational expression whose evolution through time is prominent (the trajector or primary figure). Like verbs, adverbs, prepositions, and adjectives are also relational expressions; however, in their cases, temporal evolution is less salient in the relationships profiled, being a secondary figure, or a landmark. Therefore, prepositions, adverbs, and adjectives are said to be atemporal.

Looking closely at locative expressions, Langacker (1991) shows that the verbal relationship or [PROCESS] figures as their trajectors or autonomous entities. For example, in *Mary reads slowly*, *slowly* is a type of reading. The trajector of *slowly* is elaborated by *reads*, the entity that is the profile determinant. Its landmark is any point in the scale below the normal pace of reading. In other words, an adverb is a modifier whose head (*reads*) profiles a [RELATION] as opposed to a [THING]. In psychological terms, locative expressions invoke cognitive domains and specify the relative position of the focused participants and events within those domains.

In brief, according to CG (Langacker, 1991), a linguistic expression profiles either a thing or a relationship and the nature of its profile will determine its basic grammatical class. Figure 1 shows the symbolic structures for the basic grammatical classes. A thing is

represented by a circle, a relationship by a dashed line connecting the trajector to the landmark, heavy lines indicate profiling and the heavy-line portion of the arrow the evolution through time that is salient in temporal relations.

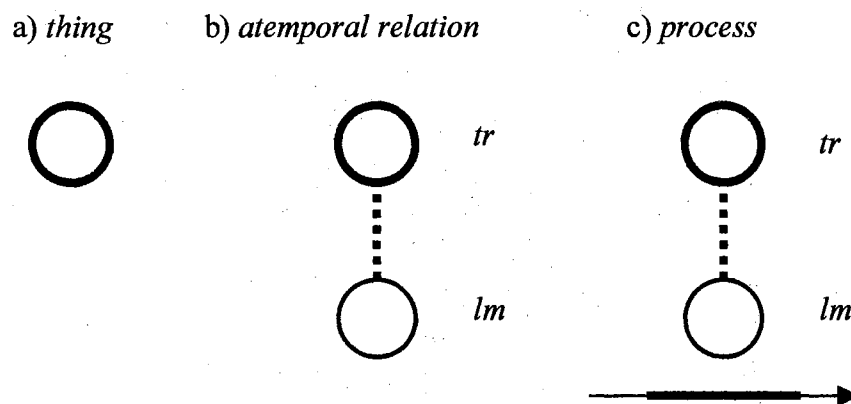


Figure 1: Schematic representation of basic grammatical categories according to CG (Langacker, 1995, p. 54).

Those symbolic structures combine to form more complex composite structures which in turn may combine to form even higher structures. Calling on these higher-order schematic notations, I will use the notation of CG to describe grammatical expression of relevance to this dissertation, beginning with word order.

Word order. Cognitive Grammar approaches word order as a schematic configuration that reflects alternative profiling options of the same image-schema as well as how we choose to segment the world for communication purposes. Different orders

profile (or designate) different sub-regions as prominent or as the trajector, yielding related but different meanings.

To clarify that understanding of word order, we may draw on Langacker's (1987) characterization of transitivity, finite clauses, and grammatical relations in general. He structures experience as an exchange of energy that varies in degree. To do that, Langacker maps Fillmore's (1968) classic definitions of semantic cases (verbs assign an agentive, instrumental, experience, goal, locative or an objective role to each linguistic expression) against schematic configurations like the ones figure 1 illustrates. Instead of talking about sentences, he talks about *prototypical events* in which discrete objects, the *trajector* and the *landmark*, interact energetically through physical contact. While the trajector is the most prominent participant of the event and the head of the action chain, the landmark is the second most prominent participant and the tail of the action chain (the dependent entity). An example of a prototypical event or action chain would be:

As equipes de exploração perfuraram dois poços direcionais.
 The (f.pl.) crews of exploration drilled (3rd.p.pl) two wells (m.pl.)... directional (pl.)
 The exploration crews drilled two directional wells.

While *exploration crews* appears as the trajector or the source of energy, *two wells* appears as the landmark or the participant that receives the energy the autonomous entity emits. The two participants are maximally antagonistic in that one is physically active and autonomous and the other is passive and dependent. Because of the energy the trajector emits, the landmark undergoes a change of state as profiled by the verb *to drill* and elaborated by the e-site *directional*. Figure 2 diagrams a prototypical action chain as conceived by Langacker (1991). The transmission of energy is depicted by a double

arrow whose direction distinguishes the agent and patient participants. The squiggly arrow indicates the patients' resulting change of state. The "v" indicates the conceptualizer's vantage point, external to the setting. That schematization also symbolizes the canonical word order in English (SVO).

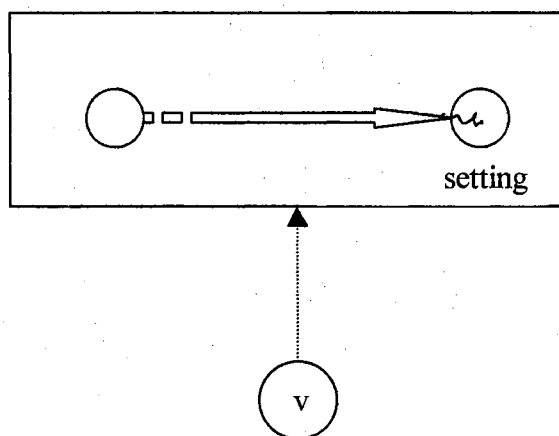


Figure 2. Word order is an [ENERGY CHAIN] metaphor (from Langacker, 1991, p.211)

Different orders relate to the same image-schema. However, each of them profiles a specific configuration. In the sentence

As equipes exploratórias passam vários dias no mar
 The (f.pl.) crews exploration (f.pl.) spend (3rd.p.pl.) several (m.pl.) days in the (m.s.) sea (m.s.)
 Exploration crews spend several days in the sea.

The conceptually autonomous structure *exploration crews* is primed as soon as one reads the sentence and remains available for further elaboration by a dependent

structure—*spend several days in the sea*. That dependent structure profiles a relation between an abstract domain (= *spend*) and the elaboration site *in the sea* that is part of it and profile determinant. Now, though the two participants—*exploration crews* and *several days in the sea* are interacting energetically, the event in which they participate cannot be considered prototypical. While one of the participants is human, carries the physical activity, and is, therefore, the head of the action (= *exploration crews*), the other is an inanimate, stative participant (= *several days in the sea*) that functions as the setting for the process the verb profiles. In other words, the two participants are not maximally antagonistic, configuring a more marginal instantiation of the action chain.

This account of sentences as events or action chains applies well to explanations of different word order in discourse and seems more adequate than all-or-nothing interpretations such as those proposed by traditional grammarians. Within CG, word order is a polysemous category that chains meaning through the trajector-landmark alignment or the head and tail metaphor. This model provides us with the theoretical basis to understand why different word orders yield similar, appropriate meanings that, abstractly and in different degrees, relate to an image-schema such as the action chain metaphor. It also provides explanatory adequacy to variations that exist in free order languages such as Portuguese in which the order of elements becomes grammatically irrelevant when in competition with morphological cues, pragmatic or cognitive considerations. In Chapter V I shall come back to this point.

Nominalizations. Nominalizations are complex symbolic units that force a conceptual reification of components of the verb schema [PROCESS] with components of the [NOUN] schema, construing the process profiled by the verb as a region in an abstract

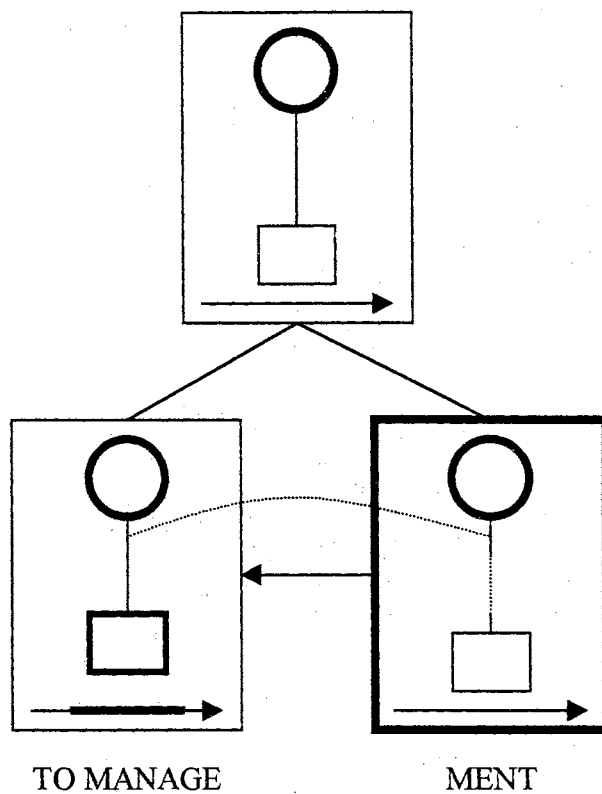
domain (Langacker, 1991). For instance, the verb *to arrive* when the suffix *-al* is added becomes the nominal *arrival* or the verb *to manage* when suffix *-ment* is added becomes the nominal *management*. Figure 3a illustrates the reification process. Because *-ment* is profile determinant, when it is superimposed on the entity *to manage*, it profiles the trajector of the composite structure [MANAGE-MENT]. Figure 3b provides the notational representation of the nominalized form.

Linguistic expressions construed as in Figure 3b are not prototypical nouns, but a marginal member of that category. To be a prototypical member of the [NOUN] image-schema, the form must be a symbolic structure that designates a region in some domain (see above) and bear the morphological trappings of the class, a characteristic that is language specific. In English, for example, nouns take inflection for number. Syntactically, the noun must enter possessive constructions and head-modifier sequences. Furthermore, a typical noun does not change its identity over time (see Taylor, 1989). The verbs/nouns “swim” and “search” are examples of nouns that change identities in discourse and are therefore less prototypical.

Prototypically, a nominalization occupies a NP position or the position of a trajector in the action-chain metaphor. That is the case of *improvement* in *The improvement of gas emission by Petrobras pleased public opinion*. However, *the improvement* is not the archetypal trajector in that it is hard to configure it as the source of energy. To construe the same event in a prototypical way, I need to place the source of energy in the trajector position and the temporal process profiling the change in the landmark (e.g. *Petrobras improved the control of gas emission*).

(3a)

MANAGE - MENT



(3b)

VERB

NOMINALIZATION

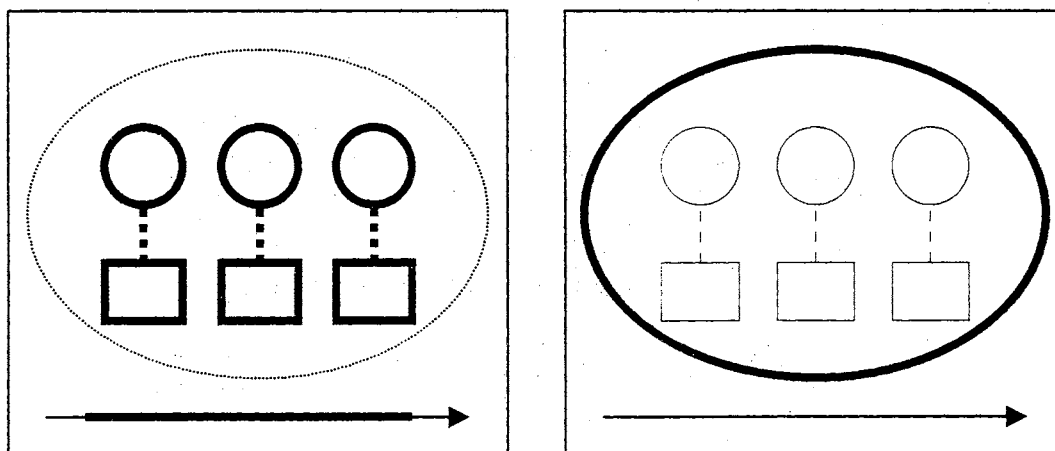


Figure 3. Nominalization according to the notation of CG
(from Langacker, 1991, p. 99 and 129)

Grammatical markings. In cognitive grammar, every grammatical structure is symbolic, including morphological trappings. For example, Langacker (1991) demonstrates that number and gender agreement are meaningful and motivated and that they display a variety of related senses describable as a network. The plural marker *-s*, characterizes internal multiplicity and cognitive salience. “Prototypically, the plural designates a mass consisting of discrete entities all of which are instances of the same type, such that any one of those entities can be individually described by the singular stem” (p.304). Instances like glasses and trousers are not prototypically plural; however, they comprise extended values of that prototype.

Similarly, Langacker (1991) points out that gender assignment may be considered instantiations of a central form that is neutral. Take the case of animates such as *irmão* (=brother), *garoto* (=boy), *cachorro* (= male dog) in Portuguese. They may be analyzed as schematic nouns with respect to the type of creature and the suffixes *-o* and *-a* as noun-forming suffixes or simply, schematic nouns themselves, their schematic meanings being [MALE/*-o*] and [FEMALE/*-a*] respectively. The integration of the two schematic nouns configures the composite structure [MALE DOG] that cannot be understood apart from its components. In the case of inanimate nouns like *braço* (=arm), *cadeira* (=chair), or *livro* (=book) the endings are equally analyzed as schematic for the noun class, being equivalent to the noun class schema: [THING/*o*]; [THING/*-a*]. Thus, the endings *-o* and *-a* are polysemous just like other grammatical constructions, and each of them has a prototypical instantiation. Now, if inanimate nouns like *braço*, *cadeira*, or *livro* will take *-o* or *-a* is a matter of usage or linguistic convention. It is part of the cognitive grammar of Portuguese

speakers. The same holds true for the fact that *livr* does not occur independently and that *-a* cannot be associated with it bearing the same meaning (*livra* is the 3rd. person singular, indicative, of the verb *to get rid of*). “Speakers master the full configuration” (Langacker, p.307).

Like gender and number, agreement markings may also be represented through symbolic units. They are interpreted as predications since their selection is determined and reflects an aspect of the argument. Take the expression *livro caro* (expensive book). In its categorizing structure [EXPENSIVE (MALE THING)/ expensive book], the trajector (=the thing) and the landmark (=the atemporal relation) are related. That correspondence is established by means of the gender marker *-o*. The marker is determined by *book* and reflects an aspect of it [MALE]. Both the trajector and the e-site contribute its own meaning, being equally salient. If that meaning is purely a grammatical function as in the case of *livro*, that is absolutely consistent with being motivated.

In an action chain like (1) the composite schematic structure results from the integration and correspondences among the schema of *-o*, the composite structure of *oil* (=óleo), the profile determinant composite structure of *black* (=negro), and the composite structure of the temporal relation profiled by *sprout out* (=jorrou).

(1) <i>O</i> (s.m.)	<i>óleo</i> (s.m.)	<i>negro</i> (s.m)	<i>jorrou</i> (3rd.p.s.; indicative mood; preterit)
The	oil	black	sprout out
the black oil sprout out			

The profile of *-o* in both structures and in the definite article that precedes the expression *óleo negro* serves to cognitively identify the black entity as a [THING] of some kind. The article, a grounding predication that indicates construal, makes salient the

[THING] that it grounds and that is its e-site [OIL]. The lack of plural marker in the trajector, the landmark, and the relational predication makes the relation between the trajector *óleo* and the profile determinant entity *negro* more salient in that it designates a mass single entity [BLACK(THING)]. That relation is further elaborated by the verb trapping *-ou* in the relational predicate [RELATION (TEMPORAL PROCESS)/...ou]. The ending *-ou* profiles a sequential relation bounded in conceived past time between the two single entities of the action chain. The final composite structure is

[THE-m-BLACK-m-OIL-m- SPROUT OUT-PAST PERF SG/ *O óleo negro jorrou*]

To arrive at it, users follow a cognitive routine that involves the superposition of the schematic structures I described.

The last grammatical expressions I would like to bring to light within a CG notational paradigm are head-modifier sequences. Take the case of the expression “black oil” (=óleo negro). While [BLACK] is a conceptually dependent predication, [OLEO] is a conceptually autonomous entity being elaborated by [BLACK] both in Portuguese and English. The difference between the language specific instantiations of the image-schema prompted by the expression is one of word order or construal and implies different cognitive routines as much as varying amounts of processing effort. Whereas in Portuguese the autonomous and more salient entity comes first, allowing the conceptualizer to construe the schematic structure of the [THING] that the writer is talking about, in English that is the last component structure an English language user configures. Combining this understanding with findings of research in perception and

information processing (see Chapter IV, Shridar, 1989), I suggest that such structural characteristic leads to differences in the time necessary to form composite schematic images and comprehend.

In conclusion, Cognitive Grammar assumes that all linguistic information is symbolic, motivated, and grounded in experience. It also assumes that language users structure linguistic information conceptually; categorize one structure based on another; conceive a situation abstractly at various levels; link aspects of different structures or elements in different domains; group simpler structures into more complex ones; impose figure/ground organization on a scene; and construe an event from different perspectives by designating unique prominence to different sub-structures. (Langacker, 1991, p.291). All assumptions are brought to bear in linguistic analysis. In addition, they constitute processing steps that users may go through to compose whole gestalts of language phenomena, being powerful tools for crosslinguistic research. A CG description allows analysts to observe what is conceptually the same and what takes different cognitive routines. This dissertation will draw on this notation and the assumptions it brings to bear.

Linguistic relativism and the cognitive approach

The rationale for including a discussion of linguistic relativism in this dissertation rests on three facts. First, this dissertation investigates Brazilian Portuguese and English discourses, two different linguistic systems; second, it tries to relate processing effort to grammatical devices that typify written discourse in the two languages; and third, it aims at proposing an image-schema for discourse in each language condition and category formation is a cultural bound phenomena. These three facts justify discussing how the

study fits within the Sapir-Whorf hypothesis or the theory of linguistic relativism—the relation between language, culture, and the mind.

In the light of cognitive linguistics, cognitive theories, and numerous crosslinguistic studies, I argue that the Whorfian hypothesis is about language performance rather than a cognitive and linguistic hypothesis about competence. This argument follows Hunt and Agnoli (1991) who interpret the Hypothesis as a psychological model for discourse analysis. According to them, writers use certain strategies because of cognitive constraints in producing and comprehending texts and, I add, because of the operationalization of the Cooperative Principle and, ultimately, the Principle of Relevance.

Some syntactic, semantic, pragmatic, and rhetorical differences across languages provide good evidence for a psychological interpretation of the Hypothesis. For example, word order is a major cue to syntactical structure and meaning in English whereas in Italian it is not (Bates and MacWhinney, 1982; see crosslinguistic studies on sentence processing in Chapter IV). Italians, for their part, rely on subject-verb agreement and animacy to retrieve meaning. Other languages like Chinese, Malay, and Japanese rely on context (Hopper and Thompson, 1993) and on the Cooperative Principle.

Another evidence for this vantage point is the categorization of the color domain. Different languages name different number of colors and in a different denotational range (for example, Danis only have two color terms), even though speakers may equally perceive the whole color spectrum (see Rosch, 1973). While facts such as color perception reflect that which is common across languages, facts related to performance such as color categorization reflect that which is different.

In still another instance, variants of the same language may have different denotational ranges for certain forms. In Isletan English some morphemes have distinctive functions. For example, the *-ing* morpheme stands for the notion of a general truth with no temporal restriction; the verb *to go* has a distributive function, carrying the notion of recurrent events. The Isletan people's acculturation process or how they structure their life experiences largely explains the different images or domains associated with these markers. Though the language the same—English—there are variations brought to bear because of acculturation processes.

A final evidence for a psychological interpretation of the Whorfian Hypothesis is rhetorical pattern. For example, Ute English writers have different assumptions about how a text should be organized and the type of information that should be included (see also my discussion of texts as linguistic, social, and cultural phenomena in Chapter II). Like Japanese or Malay writers, they expect the reader to fill in the gaps based on contextual information or shared knowledge. Anything different is a violation of Grice's maxims of quantity and relevance and of the Principle of Relevance. The same is true of Athabaskan English writers (Leap, 1989) who only include information that is not in the contexts of the text.

The differences in the way Danis code colors, Isletans code the "ing" and the "to go" forms, Athabaskans and Ute people code texts as well as the similarities among these processes illustrate how the cognitive paradigm interprets linguistic relativity.

First of all, cognitive linguistics interprets differences across languages as recurrent processes of meaning extension from a prototype. Even though alternative meaning realizations are more or less discrete, they relate to each other in a family resemblance

category. That is, they all converge to a reference point, a core that is common across languages and that Langacker (1987) calls basic domains—SPACE, TIME, EMOTIONS, SENSATIONS, ACTION, STATE. However, members of a same category relate to each other indirectly, by chaining. That is what makes translation and crosslinguistic communication possible (see my review of Toury, 1991 and Gentzler, 1993).

Consequently, though Cora grammar has many locative devices (Casad, 1987) and differs highly from English in the way it conceptualizes spatial location, we can still understand the contrast of the morphemes *u* and *a* (roughly *inside/outside*). These morphemes appear idiosyncratically in the language because of the notion of containment in physical space. The other meanings of the morphemes chain from the core, and are highlighted by the salience that the shape of the hills (top, slope, and bottom) plays in the Cora people conceptual system (they live in the mountains of Mexico). Though the structures of English and Cora are different, and the meanings not equivalent, comprehension is possible because the linguistic realizations chain back to a common domain: SPACE LOCATION.

The same is true of different forms of address in English, French, Polish and Russian, as discussed by Wierzbicka (1989). Though *Seigneur, Mr., Sir, Lord, Monsieur, pan* differ in meaning systematically, they have a prototypical semantic structure, or the same frame within a domain. Given the pragmatic contexts in which they occur across languages, we can explain this frame with semantic primitives like children, adult, people, know, familiarity, distance, respect which recur widely in the languages of the world as separate lexical items. The form *Tu* in Russian, is used by strangers or partial strangers, or yet non-intimate conversants, while in French it has a wider denotational range, applying

to children and well known interlocutors. In other words, the address forms are polysemous, forming a family of related meanings.

Second, cognitive linguistics also interprets linguistic relativity as differences in the way humans use their imaginative capacity through processes of meaning relatedness, namely, metaphors or metonymies and non-universal, socially-constructed concepts (Lakoff, 1990). What varies then is USE not concepts. Across languages, as Lakoff suggests, one is very likely to find metaphorical concepts based on universal experiences like MORE is UP, LESS is DOWN, ANGER is HEAT and so on. However, how these concepts are organized and put to use should vary substantially given the role of individual creativity, environment, and perception.

Finally, cognitive linguistics draws on notions developed by Cognitive Grammar (Langacker, 1995), advocating that different peoples structure the same conceptual scene in different ways because words, word order, grammatical devices profile different perspectivizations or paths (Casad, 1995; MacLaury, 1995). For example, while people who speak English assign secondary prominence to the object in a scene, Cora speakers see objects as energetic instigators, placing salience upon them. "In any construal of a scene, certain components are foregrounded while others serve as reference points for the characterization of the foreground" (Taylor, 1995, p.5). That explain why the sentences such as *John bought a car* and a *A car was bought by John* are not truth-conditionally equivalent. Rather, they express different "mental routes" (Taylor) that a language user takes in describing an event.

This interpretation of foreground brings to the forum facts that cognitive linguistics interprets as commonalities across languages. For example, Shridar (1989) shows that the

cognitive principles of perception and salience mediate crucial aspects of human languages such as word order, transitivity, referential expressions, and negation. While experimenting with subjects in 10 different cultural environments and language backgrounds, he found that “the order of clauses expressing perceptual events corresponds to the sequence of events in perception” (p. 223). Entities that enjoy greater salience (because of size, focus, or humanness, in that order) normally appear sentence initially in SVO languages and lead to topicalized objects; ACTIONS tend to be expressed more often than STATES; and objects on top of a vertical array are normally located with reference to that below. Therefore, perception seems to play a fundamental role in the structuring of human languages (see also Verhagen, 1995).

McCleary's (1982) investigation of the relationship between sequentiality of events, the discourse function *foreground*, and transitivity reinforces that view (see details of her methodology in Chapter II, under studies that addressed Brazilian Portuguese texts). She calls into question if the complex of syntactic and semantic features Hopper and Thompson (1980) defines as transitivity (e.g. agency, polarity, modality, participants, punctuality, volitionality, aspect or telicity, object-affectedness, object-individuation, kinesis) correlates with foreground. High transitivity has generally been seen as a means of marking certain clauses, giving them the status of figure. For example, the interplay of main clauses (figures) with subordinate clauses (ground) in discourse. However, according to McCleary's data, what is thematic in a narrative is the series of most perceptually salient events. Her subjects' judgements of whether the set of clauses in the narrative were part of the chronological sequence of events or whether they were important within the narrative reveal that chronological sequence of events and

importance are two facets of foreground. The foreground function could not subsume clauses that contained temporal sequence of events and those that subsumed the most relevant parts of the narrative in the study.

McClearly's (1982) results suggest that the general pragmatic function (foreground) that explains transitivity is a function of human cognition. Transitivity parameters such as agency, punctuality, modality, kinesis form a gestalt not because they draw attention to elements that are salient in discourse. Rather, they form a gestalt because they are attributes of a prototypical situation that is judged to be perceptually salient in the world. To conclude with McClearly's words, "transitivity owes its stability across languages to the fact that its component parts contribute to a gestalt of special significance in the human perception of experience: the prototypical causal event" (p. 73).

In conclusion, cognitive linguistic seems to offer a more comprehensive explanation about linguistic differences and similarities across cultures, building on the weak version of the Whorf hypothesis (see also Lakoff, 1990). For cognitive linguists, the differences reflect the interaction of linguistic, cognitive, and functional schemas as well as of individual creative capacities. The similarities the basic, supra-linguistic phenomena such as those related to perception. Therefore, differences among texts in different languages could be interpreted as arising from intervening variables such as the environment, knowledge structures, processing effort, and pragmatics.

Summary

This chapter described the major tenets of cognitive linguistics, and showed how it interprets categorization, prototype effects, image-schemas, differences and similarities across languages vis-à-vis studies that have examined various linguistic expressions. In

addition, it presented the notational devices of CG, applying them to nouns, verbs, locational devices, grammatical markings, and head-modifier sequences. The theoretical assumptions and expressions this chapter discussed bear direct relevance to this dissertation. Furthermore, the terminology used to describe the examples is equivalent to that of this dissertation. Last, this chapter discussed linguistic relativism in the light of cognitive linguistics. The findings and concepts the chapter brought together indicate that languages differ because variables that are not linguistic per se influence the way languages organize and use linguistic attributes. Processing effort, the environment, and pragmatics constitute three of those variables. Cognitive Grammar reinforces this psychopragmatic model of linguistic performance. It describes linguistic phenomena as the conceptualization of separate entities that need to be superposed to construe a final composite structure. This process follows a cognitive route determined by that which is linguistic per se.

CHAPTER IV
DISCOURSE PROCESSING

Overview

This chapter treats theories of attention, working memory, and discourse processing models that may shed light on possible links between linguistic devices and the effort to process information. Those theories may also help us to understand why performance limitations or enhancement occur as it is related to comprehending written information on a page. It opens with a discussion of working memory, its limitations and connection to theories of attention. Then it moves on to a discussion of different discourse processing models, perception and how it relates to comprehension across languages. The chapter also relates linguistic devices to mental processing operations, closing with some tentative conclusions and the COMMUNICATIVE TEXT image-schema.

As related to memory

Within the scope of this dissertation, the importance of memory-related studies to discourse processing hinges on two major issues: how discourse structure that is entrenched favors comprehension; and how working memory relates to the human capacity of processing information. I will first address the significance of patterns of structure for comprehension and recall of textual information. Then I will treat issues related to the limited capacity of working memory.

Existing research in psycholinguistics (Krulee, Fairweather, and Berquist, 1979; Kintsch and Greene, 1978; Meyer, 1975; Thorndyke, 1979), particularly crosslinguistic studies on comprehension (Hinds, 1984; Eggington, 1987) reveal a close dependency between the structure of a text and recall of information, mainly in delayed conditions. Krulee, Fairweather, and Berquist, working with L1 exclusively, inform us that comprehension and recall of information is better if the organizational pattern of the text fits the reader's personal experience. Hinds reveals that retention of information in memory depends on the organizational schema in which the information is presented. Because his Japanese readers were familiar with the rhetorical pattern of the essay in the study (one specific Japanese organization structure), they performed better than his English-speaking American students in an immediate and a delayed recall task on the same essay. Eggington (1987) confirms Hinds in the delayed recall task (recall of the paragraphs one week after the day they read the information) exclusively, and reinforces research on long-term memory recall (Kintsch and Greene; Meyer; Thorndyke). His twenty-seven Korean students reproduced better information that was presented in the traditional non-linear rhetorical framework of Korean only in the delayed recall task. In the immediate recall condition (recall of information after eight minutes), his subjects recalled approximately the same amount of information from both the traditional and the linear pattern texts. The linear pattern is that used by Korean scholars who have been influenced by the preferred rhetorical pattern of academic English.

The psychological significance of text structure for comprehension has also been tested in different memory experiments by cognitive psychologists (Trabasso, Secco & van den Broek, 1984; Thorndyke, 1977; Palinscar and Brown, 1984; Keenan, Baillet, and

Brown, 1984). These studies have demonstrated that memory for texts is sensitive to causal structure. For example, Thorndyke has shown that if the organization of the text conflicts with what language users would consider ecologically valid, the recall rate is poorer. While his subjects who studied an original story with a “natural” organization recalled 85 percent of it, those who studied a scrambled version recalled only 32 %.

Other studies in cognitive psychology report the effect of schemas on memory inferences (Loftus and Palmer, 1974; Brewer and Treyens, 1981; Brewer and Nakamura, 1984 for a review). Loftus and Palmer investigated the interaction of language with later reports of car accidents in two different experiments. In the first, they found that people normally overestimate the speed of a car according to the verb, which the interviewee uses—*hit* or *smash* for example. *Smash* yielded reports of higher speed. In the second, they found that as a function of time, people add information into the accident scene, fusing the actual perception with images connected to schemas in that domain. For instance, their subjects reported having seen broken glass in the accident scene. They conclude that language or words function as labels, shifting memories toward semantic domains evoked by them. Their experiment is a good example of how schemas or previous experiences combine with language, interfering with precise recall. Later elaborations combine with the original perceptions of an event forming a new unit of memory.

Similarly, Brewer and Treyens (1981) note that because of pre-existing schemas, subjects recall attributes which are typical of the category but which fail to appear in the specific entity they are asked to describe. Brewer and Treyens’ subjects were asked to write down what they could remember about an experimental room in which they waited

for the researchers. In the description of the room, subjects included the attribute *books* that did not appear in that specific instance of an experimental room. On the other hand, the majority of Brewer and Treyens' subjects failed to include attributes which are not typical of an office, namely a bulletin board and a skull, but that were part of that office. In brief, while typical attributes of an office were well recalled, even when they were false of that particular instance, less typical attributes were less recalled. Loftus and Palmer's (1974) and Brewer and Treyens's experiments attest to the psychological reality of schemas. Whenever conceptualizers recognize an entity as being a member of a certain category, they infer values that are associated with that image-schema (Reed, 1996), enhancing the construal of meaning.

Schemas affect text comprehension not just during recall, but also during the comprehension process that precedes recall. One source of evidence for that assertion is Smith and Swinney (1992). They taped forty adult subjects while they read vague texts that were similar to Bransford and Johnson's (1972) "washing clothes" story. Half the texts had a title that activated a relevant schema and the other half had no titles or no schemas. The presence of schemas speeded reading times for 17 or 18 story lines. When no schema was available, reading times were faster for sentences with more repeated concepts (identical repetitions, identical root-lexical items, synonyms, anaphoras). In addition, the presence of both a title and concept repetition interacted significantly with reading times. Their findings suggest that bottom-up and top-down strategies combine to facilitate comprehension.

The second memory-related issue of relevance to this research is the limited capacity of working memory. Let's first discuss what working memory is and then we may address

its limitations. Working memory is a type of short-term memory (STM) that maintains and manipulates information. Nowadays, theories of STM are in disfavor because of the connectionist approach to information processing (see the discussion of models of discourse processing in the next sub-heading), among other things. Cognitive psychologists such as Anderson (1991) suggest that STM and long term memory (LTM) are simply two sides of the same coin. While STM serves to activate the nodes (items) within the network, LTM is the associative network itself. Others like Baddeley (1986, 1990), propose that STM is actually an attention mechanism to solve problems, to make decisions, to compare, to select, and so on. When you are reading you are receiving information from the linguistic environment as well as from other contexts such as your world knowledge. You need to use all sources of information to perform the task of comprehending. Therefore, short-term memory is often referred to as working memory or memory that is active at a point in time to perform a specific task.

Though there is too much about working memory that cognitive psychology still needs to unravel, Baddeley (1986; 1990) seems to offer the best explanation of how we use STM (short-term memory) as a working memory or working “attention.” According to him, we have a working memory that has a central executive and several slave or auxiliary systems that keep information available. Two of these systems are the visuo-spatial sketchpad and the articulatory loop. The executive assigns information to the auxiliary systems, retrieves it or moves it among the systems.

In addition, Baddeley (1986, 1990) proposes that what limits the length of the memory span is the speed at which the auxiliary systems rehearse information (visually or verbally) and keep it active. If we try to keep too many items in working memory, by the

time we get back to rehearse the first one, it will have faded away to the point where it takes too long to retrieve and re-rehearse it. Items that are active in working memory fade away in 20 to 30 seconds to give place to new or associated items (Reed, 1996).

Cavanagh (1972), working on the premises of a short-term memory that stored information, has paved the way to Baddeley's claims about speed of rehearsal and the limitation of the auxiliary systems. He has demonstrated that the faster the memory search rate, the greater the number of items that can be under the focus of attention.

Previous research that investigated memory span (Miller, 1956; Sternberg, 1966; 1967; Anderson, 1976; Just and Carpenter, 1987; 1992) or how much information people could recall in digit span tasks has also contributed to our understanding of working memory. Miller indicates that adults often recall a string of seven letters at a point in time and that that number may vary from five to nine items. That explains why telephone numbers have seven digits (Anderson, 1995; Reed, 1996). Just and Carpenter (1987; 1992) show that the working memory buffer for text retains no more than 2-5 clauses at a time, or approximately 8-20 seconds of verbatim text. Sternberg relates the limitations of working memory to recency of cues and proves that the time required to identify an item in working memory increases as a linear function of the number of items under the focus of attention. His research suggests that people search items in working memory one at a time. Adding to Miller and Sternberg, Anderson (1976) suggests that it is how frequently and how recently we have used a memory (its level of activation) that determines the speed and the probability of accessing it. That is, the level of activation determines both the probability of access to a memory and the rate of access (Anderson, 1995).

Those findings were further reinforced and developed by Ratcliff and McKoon (1981) in a demonstration of associative priming and its relation to spreading activation. While associative priming refers to how the presentation of an item (e.g. *doctor*), which bears some kind of similarity to a second item (e.g. *nurse; hospital; prescription; pharmacy; etc.*), facilitates the access to the second, spreading activation defines the flow of energy in a semantic network formed by those associated items. Networks of associated information are formed by members of the same category or items that bear physical or functional similarity. Ratcliff and MacKoon's results indicate that activation spreads through a network from items under the focus of attention to associated memories. The more that activation spreads to available material, the more rapidly it can be retrieved.

Finally, the amount of activation that spreads to a memory also depends on the cue strength of that memory. The more a memory is practiced, the stronger it becomes and the more successfully it can be retrieved (Pirolli and Anderson, 1985; Anderson, 1995). However, the strength of a memory trace decays as a power function of the retention interval. That is, the longer the interval between the moment you activate a memory and the time you use it, the more it decays (Bahrick, 1984; Wixted and Ebbesen, 1991). These findings reinforce Cavanagh's (1972) and Baddeley's (1986, 1990) claims about the speed of access to a memory and recall.

The role of working memory and the limitation of its auxiliary systems have also been applied to written language production. Two recent models of the writing process (Hayes, 1996; Grabe and Kaplan, 1996) emphasized its central role which was almost simultaneously sanctioned in a model of working memory in writing proposed by Kellogg

(1996). The three studies apply Baddeley's (1986) conception of working memory.

Hayes, and Grabe and Kaplan offer a socio-cognitive model in which on-line processing assembly of writing is subject to the articulatory loop. In other words, many writers plan their written sentences by vocalizing and sub-vocalizing in the medium of speech.

However, because the articulatory loop has a limited capacity (6 digits approximately), that limitation is transferred to the writing process. Furthermore, in the mentioned models of composing, cognitive processes freely interact with other systems such as the situational context (in both models) and with affective, motivational, and the visual systems (in Hayes model only). "If we are to understand how texts are understood and how they are best designed, we have to attend both their verbal and their visual features" (Hayes, p.22).

Figure 4 illustrates Haye's model and the interaction among the affect-motivational, cognitive, and environmental systems.

With regard to Kellogg's model of working memory in writing, it specifies how the formulation, execution, and monitoring phases of the writing process call on the auxiliary systems of working memory (e.g. the auditory loop and visual sketchpad). According to Kellogg, formulation demands the most from the auxiliary systems because to plan and translate ideas to the page, the visuo-spatial sketchpad, the phonological loop, and the central executive systems need to make decisions about what and how to write.

Similar to the body of research on memory span, strength, and decay, Hayes (1996), Grabe and Kaplan (1996), and Kellog (1996) relate writing production to limitations that have been typically associated with language. That suggests that production and comprehension of written language may reflect two sides of the same coin.

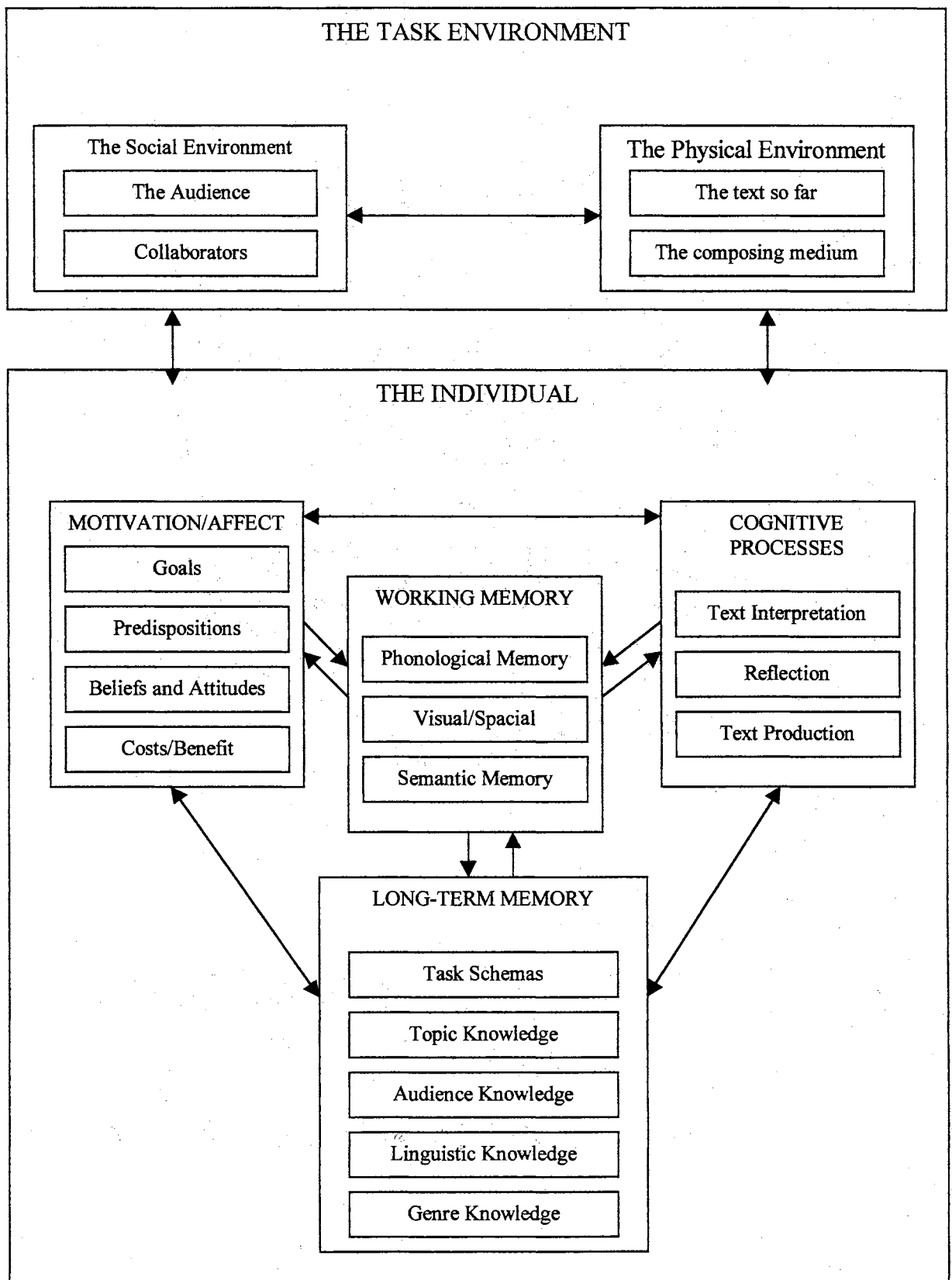


Figure 4. A socio-cognitive model of the writing process (Hayes, 1996, p.4)

If you refer to Chapter II, you will see that Chafe (1994) has been working with the same premise for some time.

Besides Chafe (1994), other discourse analysts have been applying findings of research on memory to discourse analysis. Namely, Givón (1995), Tomlin (1987) among others (see Chapter II). This dissertation follows a similar trend. It draws on the idea that the human information processing systems are limited resource systems and that both production and comprehension are subject to that limitation. The executive system assigns maximum attention to the particular part of information that is under our visual field and attenuates the resources given to parts that are out of the spotlight. A theory of working memory is a theory of allocation of resources. If the person wants to process material in other parts of the visual field, it is necessary to move the spotlight and this takes time, being thus costly (Posner, Synder, and Davidson, 1980; Erikson & St. James, 1986). In addition, to ease the effort of processing, save time, and enhance comprehension the system also maps data under the focus of attention onto existing schemas and infers values that are associated with it because of the spread of activation.

Models of discourse processing

The last ten years of research in cognitive psychology has produced models of information processing that combine bottom-up with top-down mechanisms, stressing knowledge-based strategies. One such example is Rumelhart and McClelland's (1986) parallel distributed processing model, a connectionist view to processing that sees context and the information under focus as triggers for meaning construal. In Parallel Distributed Processing (PDP) the connections among the nodes of the network hold information (see Anderson, 1976, in the discussion of discourse processing as related to memory),

spreading or inhibiting it depending on contextual effects. The connections go from every cognitive unit to every other unit within the same level and across and you can access information in memory based on nearly any attribute of the representation you are perceiving.

Take the case of language processing in which syntax, semantics, as well as other factors, including world knowledge and schemas conspire to guide comprehension. If I say *I saw the Niagara Falls flying to Montreal* your encyclopedic knowledge tells you that *The Falls* cannot fly. That hidden cognitive unit activates the one possible version that makes sense and inhibits others that do not. In many other instances, schemas determine interpretation. Language users recall attributes that are typical of the categories under focus (in this case *I* and *the Falls*) and combine them with other concepts present in the text (*saw* and *flying to Montreal*). While *I* is the typical source of energy that can see and fly, *Niagara Falls* is less typical. Therefore, the SOURCE OF ENERGY schema inhibits other possible interpretations in that *I* fits in the typical action chain metaphor as the trajector. Another example is the Warren's (1970) phonemic restoration effect. When any type of noise replaces a sound in a word (e.g. *tel...vision*), people fill in the missing sound and process the whole word. The activation of neighbor sounds or lateral activation in the neural net conspires to strengthen successful processing of information.

In brief, Rumelhart and McClelland (1986) proposed a distributed superpositional approach to information processing. In such a model, memory consists of traces resulting from specific experiences, and generalizations emerge from the superimposition of the specific memory traces. We do not keep each trace in a separate place, but, rather, we superimpose them so that what the memory contains is a composite (Reed, 1996).

However, the connectionist approach fails to mention how we select the traces we superimpose and allocate memory resources adequately.

To accommodate existing data about language, including the limitations of auxiliary systems that deal with memory traces, research on discourse processing has combined Baddeley's (1986, 1990) model of working memory with Rumelhart and McClelland's (1986) PDP model.

The first model to apply the concept of networks and working memory to discourse processing (narratives, specifically) was the explicit textbase model (Kintsch and van Dijk, 1978; van Dijk and Kintsch, 1983; Kintsch, 1988). Because the model is fairly complex, I will summarize its major assumptions. There are two input units in the model: the reader and the text. While the knowledge and goals of readers influence what is relevant for comprehension and guides their inferencing of hidden cognitive units, propositions (idea units) divide the text into meaningful units and form a semantic network. The task of comprehenders is to construe a mental representation of the information provided by the input. They do that in cycles (propositions, paragraphs, texts) because of the limited capacity of working memory.

According to Kintsch (1979), only an estimated number of four propositions can be kept active in the network given the limited capacity of attention mechanisms in working memory. Consequently, whenever incoming propositions can be integrated with information readers have already encountered (semi-active), the processing of information demands less effort. The further back in the text the unit is, the more time it takes to superimpose that unit onto other current memories (see my review of discourse processing as related to memory). When the proposition under the focus of attention overlaps with

ideas that are still active in working memory, comprehension is at its best. However, if readers must first search for an overlapping idea in the network to reinstate past propositions and then form, with current information, a composite structure, processing slows down, becoming more difficult. Furthermore, if comprehenders fail in their reinstatement searches because the associated expression is no longer active, they must start a new network, failing to integrate ideas successfully.

Another assumption of the model that ties back to research on memory (see above) is that the longer a proposition is held active, the greater its cue strength and its eventual recall (see Anderson, 1995).

Van Dijk and Kintsch (1983) added to the 1978 explicit textbase model proposing the Construction-Integration model. In it, a situation model (lexical and syntactic information in the text combined with prior knowledge) integrates both top-down and bottom-up processes, following the construction phase that relies on bottom-up activation of information in the semantic network. The situation model facilitates reinstatement searches and inferencing processes, and enlarges the capacity of working memory. Thus, now we have three cognitive units: a mental representation of the text itself (the textbase), a representation of the situation described by the narrative and integrated into the reader's previous knowledge (the situational model), and the propositional structure (data-driven). The structure of the situation model reflects primarily the structure of the situation described by the story, being a domain structure independent of the rhetorical organization of the narrative (Kintsch, 1995).

This model has been applied by different researchers and has yielded a consistent corpus of evidence that supports it (Weaver and Kintsch, 1991; Kintsch, 1994; Barsalou,

1992) as well as a number of studies on the readability of texts (Kintsch and Vipond, 1979; Kintsch, 1979; Fletcher, 1986; Britton, Van Dusen, Glynn, and Hemphill, 1990; McKoon and Ratcliff, 1990; Swinney and Osterhout, 1990). Among other things, readability studies indicate that word frequency, the number of reinstatement searches and sentence length are the best predictors of readability, if readability is defined as the number of recalled propositions divided by reading time. Another predictor is the number of inferences readers need to make to construe meaning. Whenever a concept is not directly repeated, readers need to infer it from the connections that are active. This demands time and delays the comprehension process.

Accounts of discourse comprehension within the van Dijk and Kintsch's Construction-Integration Theory make strong claims about data-driven (the lexicon and syntax) mechanisms only during the construction phase. They portray those mechanisms as simple aids to how a reader should parse a text into meaning units (propositions) and construe the situation model (Kintsch, 1995; Grabe, 1996). The backbone of the process, however, is the hidden units: background knowledge and inferencing.

Such understanding is not surprising since most of those accounts dealt with narratives, a type of text that, according to research (Britton, 1994; Graesser and Britton, 1996; Lorch, 1995), demands more background knowledge and inferencing than expository texts. While narratives have a goal-outcome macrostructure and a linear microstructure based on causal relations, expository texts are hierarchically organized around topics, invoking few inferences (Lorch, 1995). In expository texts, therefore, linguistic cues become the primary factors in the construction of local coherence and, ultimately, of a text mental model. Though different languages require different amounts

of inferencing depending on the relevant cultural model (see Chapter II), the informational dimension of expository texts seems consistent across languages. Paraphrasing Britton (1994), if the writer puts the correct instructions on paper and readers properly use them, they get the structure of ideas. If the writer, on the other hand, puts wrong instructions or fails to insert them in sentences, readers may get wrong, incomplete or no idea structure at all.

Taking that evidence into account, more recent models of discourse processing (Kintsch, 1995; Gernsbacher, 1996; Goldman and Varna 1995; Perfetti, 1994; Graesser and Zwaan, 1995), have moved the focus away from knowledge based processes to data-driven mechanisms in discourse comprehension. Kintsch (1995) re-addresses the issue, re-examining how readers construct situation models for stories. Actually, Kintsch incorporates Givon's (1989; 1995) proposal that the grammar tells the reader quite precisely where to look for what in a text.

In the light of that understanding, Kintsch (1995) re-studies Gernsbacher and Shroyer's (1989) investigation of discourse fragments in narratives, Trabasso and van de Broek's (1985) analysis of causal reasoning in story understanding, and Kintsch's (1976) macrostructures and summaries. His findings led him to affirm that syntax together with some weak semantic principles allows readers to generate situation models that are quite adequate regardless of background knowledge or inferences. He shows, for example, that a single verbal signal may determine if a text representation will be organized linearly or in terms of a topic-sub-topic structure. He also analyzes how syntactic cues indicate causal relations in a text and shows that syntactic cues yield the same recall pattern a strong knowledge-based model would predict. If we take into account syntactic signals, precise

causal inferences are not necessary to predict either recall or summarization. Therefore, like perception, comprehension is highly data-driven. If difficulties arise, more controlled problem-solving behaviors take place. In his words “If we attend to the processing instructions the syntax gives us, our model of language comprehension will be much more powerful and capable of dealing with a range of phenomena that were previously out of its reach” (p.158).

Grabe (1996) reviews different models of comprehension (McKoon and Ratcliff, 1992, 1995; Fletcher, 1994 among others) in which knowledge-based mechanisms are major contributors to coherence of narratives. He corroborates Kintsch (1995) and affirms that the data available does not support a strong top-down view of discourse processing. In all models he reviewed, linguistic cues directed the comprehension of narratives as well.

Compared to Kintsch (1995) and the models reviewed by Grabe (1996), Gernsbacher (1996) proposes a more strongly data-driven model—the Structure Building Framework model—that applies to an array of genres and the two language modalities. In her model, the linguistic systems establish the foundation of discourse processing and direct toward coherent interpretation. The mechanisms involved are processes of structure building, mapping and shifting, and processes of enhancement and suppression (activation or inhibition). Structure building operations are triggered by linguistic cues in the text and generate a network that strengthens or inhibits the connections among units (see also Britton’s analysis of expository prose and the role of structure building operations in comprehension, 1994). According to Gernsbacher’s model, most problems of comprehension derive from readers’ inability to suppress unimportant information

efficiently. That inability overloads working memory capacity and the ability to map active information onto units farther back in discourse.

Challenging Gernsbacher (1996), Whitney, Budd, Bramucci and Crane (1995) dispute the assumption that knowledge-based mechanisms have outlived their usefulness. They reviewed existing discourse processing models and conclude that none of them adequately addresses the variability of top-down processes across contexts. According to their cross-examination, the role of top-down mechanisms in comprehension is a function of the reader and of the context. Thus, they propose a new framework that focuses on how the comprehension systems distribute and maintain activation throughout discourse (The Distributed Activation Control framework). Within the DAC the management of excitatory and inhibitory connections among concepts of the network originate comprehension. In other words, the same processes that account for context effects on lexical access tasks guide which information will be kept in working memory and which will be fired. Top-down processes remain important.

To summarize, existing models of discourse processing fall in a continuum from those that strongly favor top-down processes to those that strongly favor bottom-up processes and the limitations of working memory. Though the debate is still on, there is mounting evidence that background knowledge and inferences face the competition of linguistic systems as strong keys to discourse comprehension. Linguistic systems may not rule out the situation model argument. However, they certainly reinforce it (Grabe, 1996).

The next section addresses the type of instructions linguistic devices send to the processor in sentence and discourse comprehension. In addition, because of the scope of

this dissertation and of the wealth of information morphology systems carry (see Chapter three), the next section emphasizes findings of crosslinguistic research.

As related to linguistic devices

Various research studies in cognitive psychology have demonstrated that proper names (Garrod and Sanford, 1990; Vonk, Hustinx, and Simons, 1992), temporal and spatial expressions (Givón 1995; Sanford and Henderson, 1982; Anderson, Garrod, and Sanford, 1983), quantifiers (Moxey and Sanford, 1993), nominal reference, pronouns, full definite NPs, indefinite NPs, restricted relative clauses (Garrod and Sanford, 1985; Tomlin, 1987; Givón, 1995), and other linguistic devices instruct readers to superimpose, re-activate, activate or eliminate connections of the network toward the construal of a textual mental model. While proper names serve to activate a node that will persist throughout discourse as a main character, a full noun phrase or role description activates a node that will persist throughout discourse as a secondary character.

For example, Givón (1995) proposed a cognitive model in which grammar is a mechanism for speeding up the processing of both local and global aspects of text coherence. Devices such as agreement, same-subject cataphoric morphology, restrictive modifiers, and word order instruct the processor to activate, continue the activation or terminate the activation of a node in the network. While agreement and cataphoric morphology signal continued activation of the same node in the network (the topic referent), restrictive modifiers, relative clauses and word-order devices such as left-dislocation terminate the activation, exhilarating a new node.

Specifically, Givón's (1995) investigation of coherence indicates that "anaphoric connections are time-saving devices in the processing of coherent discourse" (p.105).

They ground incoming new information to some pre-existing composite structure in the text or in the readers' mental entities. That is true of zero anaphora and unstressed pronouns, devices that cue maximal topic continuity or connection activation of the same node in the network. On the other hand, stressed pronouns and full lexical nouns signal topic discontinuity or de-activation of a node. When readers run across a full indefinite NP or a stressed pronoun in discourse, they terminate the activation of the preceding chain, and open a new file or node as the topical referent. Whenever a connection is terminated, the processor needs to activate a new node. This takes time.

Furthermore, his research indicates *this*-marked NPs cataphorically ground indefinite referents, activating a node that will be from then on accessible for further associations. Restrictive relative clauses serve the same grounding function. They tag the newly-introduced NP as an important topic in subsequent discourse, making it perceptually salient and ready for continuity. Even punctuation functions as cataphoric grounding, flagging the degree of thematic continuity of the next clause.

Givón (1995) also illustrates how temporal and spatial expressions signal an episode or theme shift (see also Anderson, Garrod and Sanford, 1983; Sanford and Henderson, 1982) when sentence-initial or when opening a paragraph. As he says, sentence initial adverbials of time and place are better described as theme switching devices or in Fauconnier's (1994) terms triggers that set up a new mental space in which discourse will unfold. Temporal and spatial expressions move the focus of attention from the action itself to themselves. Even when there is referential continuity, thematic discontinuity overrides it to set up a new domain. If a given mental space is salient within a cultural or general experiential framework, it is going to be unmarked in discourse or naturally

activated. However, if a given space builder is explicitly mentioned in discourse, the fact that something is said makes it cognitively salient or prominent.

In brief, for Givón (1995), grammar, specially morphemes and syntactic constructions, are sets of mental processing instructions designed to trigger specific mental operations. These mental operations involve (a) attention activation and (b) search and retrieval in an episodic memory of the text represented as a network of connected nodes. The nodes may be connected either sequentially or hierarchically to other nodes in the text-structure and the more connections they have the more grounded or mentally accessible they are.

Tomlin (1987) reinforces Givón's (1995) interpretation of linguistic expressions as cognitive events or mechanisms that allocate attention during the on-line process of discourse. He differs from Givón in that he analyzes discourse from the perspective of production and not comprehension. Focusing specifically on the syntax of reference in the on-line production of narratives for a slide picture sequence and a videotaped cartoon, he demonstrates that subjects use a full noun to reinstate reference across an episode boundary and a pronoun to maintain reference within a particular episode. The recency/distance approach (Givón 1983; to assess referential distance we count the number of clauses that intervene between a referent and its last antecedent) failed to fit his experimental data. Subjects used full nouns to refer to an antecedent only a single clause away even in cases in which ambiguity resolution was not at issue. Also, subjects sustained pronominal reference for more than a clause or two. He concludes that the alternation between noun and pronoun is a function of the limited capacity of working memory, which is manifested in discourse through episodic organization (thematic

organization). Linguistic codes facilitate the cognitive processes required of the listener to build the intended mental representation of discourse. While some syntactic constructions facilitate activation, others facilitate selective focusing.

Research has also investigated how linguistic devices such as nominalizations reflect cost of processing. Chafe and Danielewicz (1987) indicated that the high frequency with which nominalizations appear in discourse means they place no constraint on memory. Contrastively, Brennan and Brennan (1990) show that nominalized clauses constitute one of the major sources of misunderstandings in the courtroom. Specifically, they examined 26 transcripts from a testing program designed to identify, quantitatively and qualitatively, features of court-room language that place a great degree of difficulty for unsophisticated and young language users (their ages ranged from 6-15). Out of 13 descriptors, nominalization appeared in third place as the source of difficulty. Commonly, court-room language removes the agents and the recipients of actions, objectifying it as a thing. For example, instead of saying *Susan hit Bill's leg* (Brennan and Brennan, p.65-66), a lawyer would perhaps say *the hitting of the leg*. In that the action is objectified, there is no hint about who did what to whom. Nominalizations become a strategy to avoid inducing answers, and to make the statement impersonal, detached from reality. Brennan and Brennan cited McCawley (1970) and other studies of jury instructions that addressed nominalizations. In all of them nominalizations have appeared as sources of misinterpretation to adult witnesses as well. In the Charrows study (Danet, 1980) only 29% of the replies to questions which contained nominalizations were correct. Those instructions that did not contain nominal forms received 40% correct answers, a fact that indicates that comprehension improved. Brennan and Brennan conclude that

nominalizations in their sample functioned as distancing devices because they led children to feel as if they were watching rather than participating in the jury procedures.

Indirectly, Halliday (1989) also suggests that nominalizations place constraints on processing. The biggest evidence, he claims, is that young children fail to understand them until the age of eight or nine. Only then do they master grammatical metaphors such as nominalizations.

Cognitive psychology has proved that point in experiments in which performance on free recall tasks interacted with word frequency (Hall, 1954; Sumbly, 1963; Saliés, 1995a). The rarer the word, the greater the effort to process it, and the lower its recall scores. This is the case of nominalizations. They are typically low frequency words that enjoy low cue strength. Therefore, they are likely to increase the cost of processing, mainly when they appear in discourse with great frequency.

This section has illustrated how morphology and syntax relate to cognitive and perceptual structure. Linguistic devices may activate, speed up or inhibit different connections in the information processing network. Anaphoric morphology is of special interest to this study in that it is a time saving device, and Brazilian Portuguese systematically makes use of it to maintain active the same referent within an episode. To fine grain our understanding of how linguistic cues operate in languages morphologically similar to BP, the next section addressed crosslinguistic research on sentence processing in Italian, French, and Spanish.

Crosslinguistic studies on sentence processing

Experimental crosslinguistic studies that have investigated linguistic cues and their role in sentence/discourse processing corroborate the view of expressions as instructions

to the processor. Furthermore, those studies have demonstrated that the more ecologically valid cues like word order, animacy, agreement are, the more likely it is that they will influence processing strategies in a specific language (Bates and MacWhinney, 1989; Slobin and Bever, 1982; Kail, 1989; Charvillat and Kail, 1991; Aaronson and Ferres, 1986; Hoover, 1992; Cuetos and Mitchell, 1988; Hunt and Agnoli, 1991; Shridar, 1989; Flores D' Arcais, 1987). The term ecological validity refers to frequency of use, salience, entrenchment in a given psychological reality.

Among the first to conduct studies of this kind were Slobin and Bever, who, in 1982, published a seminal work on the crosslinguistic study of word order and inflections in English, Italian, Turkish, and Serbo-Croatian. Their study involved 48 subjects for each language. Results show that from an early age children construct a canonical sentence schema, derived from linguistic experience, that embodies the typical attributes of sentences in their respective languages and that serves as a central category for the application of productive and perceptual strategies. Whenever the canonical sentence or distinct word order and inflectional strategies which are typical of their languages failed to appear, children failed to understand. It is likely then that the more a discourse shows canonical schemes or more typical structures, the higher the likelihood that readers will retrieve the intended image-schema the addresser has tried to convey.

Almost simultaneously, Bates, MacWhinney, and colleagues began to publish a series of studies on sentence processing in a variety of languages (see Bates and MacWhinney, 1989 for an overview). The original purpose of those studies was to collect evidence for Bates and MacWhinney's (1989) competition model of sentence processing, which supports a unified functionalist approach to language development. Manipulating

word order, animacy, morphological inflections, topicalization devices, and prosody randomly, the studies would present two nouns and a transitive verb (e.g. *kisses rock zebra*) to native speakers and ask them to choose an actor. Though the cues and the stimuli varied from study to study (some studies used visual stimuli like plastic toys, others oral stimuli) the basic idea was the same. I will review here those studies that involve English and/or Italian, French, and Spanish given the scope of this dissertation, as well as some findings about German and Hungarian because of particular differences in the methodologies of the studies (on-line vs. off-line technique in the case of German and removal of marking cues in the case of Hungarian). For studies about other languages see Kilborn and Ito, 1989, for Japanese; Sokolov, 1989, for Hebrew; MacWhinney, Pléh, and Bates, 1985, for Dutch, among others.

Let's start with English and Italian. Bates, MacWhinney, Devescovi, and Smith (1982) have investigated functional constraints on how 30 adult-native English speakers and 30 adult-Italian speakers processed sentences in their respective languages. The researchers used a set of toy plastic animals as stimuli and a set of 81 sentences, which varied in word order, animacy, topicalization and contrastive stress. Subjects had to interpret them. Their responses reveal that Italians relied more than English listeners on semantic strategies to process sentences, primarily on animacy and topicalization. For Italian subjects, the notion of "subject" was closer to topic. Italian subjects also took longer to apply word order strategies even when there was no conflict between cues. These findings suggest that, as far as processing is concerned, Italian is "less" of an SVO language than English. Systematic facts about semantics and pragmatics contribute to

syntactic processing in that language. English subjects, on the other hand, relied on word order, interpreting even second nouns in non-canonical word orders as subjects.

Similar results concerning English and Italian were found by MacWhinney, Bates, and Kliegl (1984) who also investigated German-speaking adult subjects and included an analysis of processing strategies in their study. This time, Italians relied primarily on agreement cues, and Germans on both animacy and agreement, in that order. As for processing strategies, English-speaking subjects waited for the entire sentence to make thematic assignments, adopting a late-closure strategy. Germans and Italians, for their part, favored on-line processing strategies, that is, they used agreement and animacy cues to assign roles as the cues appeared in the sentences. These strategies proved to have high cue applicability and reliability. Therefore, rich verb morphology and case marking related to processing strategies.

Findings about English and Italian speakers were again replicated by Bates, MacWhinney, Caselli, Devescovi, and Venza (1984). This time, children between the ages of 2 and 5 were the subjects. Results reveal that children, from an early age, show sensitivity to the relative information value of the various cues in their native language cues (see the 1982 Slobin and Bever's study for similar conclusions). While American children relied on word order, Italians relied on semantic cues. However, whenever they failed to appreciate the discourse functions of those cues, they also failed to make full use of interpretive cues. Those findings further reinforce the connection among semantic, pragmatic, and syntactic systems.

In addition, Mac Whinney, Pléh, and Bates (1985) conclude that ecological validity or the interaction of the environment with comprehension equally influences sentence

processing. Their investigation involved the impact of case marking, word order, animacy and other cues on the choice of an agent by Hungarian adults and children. Though for Hungarians case marking was the strongest cue at all ages and the first-noun-as-agent strategy predominated, the researchers obtained similar results after removing all case marking cues. This implies that subjects were choosing an actor based primarily on discourse functions and supra-linguistic schemas against which they were mapping discourse.

Kail (1989) shows that French speakers rely on SV agreement and cliticization to choose actors. Though the canonical SVO order predominates in frequency, word-order processing in French always interacts with cliticization agreement, depriving the canonical vs. non-canonical dichotomy of its psycholinguistic relevance (see Charvillat and Kail, 1991 for a study about word order in French). Like in Italian, semantic and morphological cues are more important than word order for sentence interpretation in French, which makes French “less” of a SVO language than English with respect to processing. Based on these findings, we should expect and indeed hope to see not only different cues operating in different languages, but also the predominance of those with high cue strength and ecological validity, for example, SV-clitic agreement and animacy in Italian.

Finally, Kilborn (1987) replicates the importance of morphology as an aid toward interpretation. He investigated German speakers using a German on-line version of the task, allowing subjects to make their decisions as quickly as possible. As soon as the relevant morphological information was available, German subjects would offer an interpretation. His results are similar to those of off-line studies.

According to the research here presented, morphology and syntax carry psychological meaning, cueing instructions to the processor. Take the case of word order. Word order seems to be a prototype-based category (Bates and MacWhinney, 1989). In French, though there is a central member (SVO) which is highly frequent and ecologically valid, other members of the word order category are equally acceptable if there is a clitic and agreement trappings guiding interpretation.

In fact, Lambrecht (1987), in an intriguing study about the status of SVO sentences in French spontaneous oral discourse, demonstrates that the SVO pattern is not the predominant pattern at the level of surface structure. His corpus revealed that the preferred clause of spoken French contains no subject NP and that French speakers employ various grammatical constructions to preserve the preferred structure (e.g. presentational constructions; left dislocation; clefts). Sentences containing a lexical subject were rare. Furthermore, he found a correlation between the occurrence of lexical subjects and the discourse status of the clauses in which they occurred. Lexical subjects strongly tended to appear in clauses that represented pragmatically backgrounded portions in the discourse. He cited François's (1974) similar results in a study in which out of 1,550 nouns present in a corpus of conversations only 46 were lexical subjects. The vast majority of nouns in the corpus appeared in prepositional and adverbial phrases, in extra-clausal topic phrases, and in phrases that had no syntactic connection with the proposition, neither in the subject nor object position. According to Lambrecht, the overwhelming frequency with which the syntactic sequence clitic-verb-NP appears in spoken French indicates its cognitive or processing advantage over other possible sequences such as the SVO.

Likewise, Brazilian Portuguese (see Chapter V), traditionally seen as a free order language, accepts with high frequency variations from the canonical SV(O) order because morphology, prosody, and pragmatics cue comprehension. Those cues are the means through which the SVO central member motivates the existence of other members, extending the category. In BP sentences with V (O) order are quite common and acceptable. Indeed, the presence of full NPs or stressed pronouns sound foreign in extended discourse and is typical of learners of Portuguese as a second language. That is not true of English, a language that lacks the richness of morphology of French or Portuguese or Italian, among other languages.

The present review has made clear that morphological, semantic, and pragmatic cues influence sentence processing originating different gradiences of membership in the SVO language category. Languages that show motivated variants of a central member of the SVO category are less SVO order than those that do not. However, it is the central cognitive model--the SVO category--that governs possible variations. As Slobin and Bever (1982) have demonstrated, language users map the variations onto an existing canonical schema to comprehend.

The present review has also indicated that languages like English, which relies primarily on word order to construe meaning, lead users to employ a late closure strategy. That is, they must hear or read all of the words to know the complete order. Then they can assign grammatical roles. On the other hand, inflected languages like Spanish, Italian, and French whose verbs and clitics carry information about the grammatical relationship among the noun phrases in general, allow an on-line processing strategy (also called distributed processing or local parsing strategy). Basically, what these studies have shown

is that processing strategies are responses to typicality of linguistic cues that, in turn, are language specific.

As related to other cognitive factors

Other crosslinguistic studies involving comprehension and metalinguistic processes also build on findings this review presented. Aaronson and Ferris (1986) asked Chinese-American bilinguals and English monolinguals to read and rate English words for their contribution to the structure and meaning of a given sentence. Their results suggest that language specific differences between Chinese and English relate to differences in the cognitive processing of linguistic information. Though the information communicated was the same in both languages, bilinguals rated English words in the various lexical categories as contributing more to both meaning and structure of their sentences than did monolinguals. What English does with individual words such as auxiliary verbs or tense, and aspect, Chinese does with sentential context. That is, Chinese is a more context-dependent language than English as much as English is a more context-dependent language than Italian or French. Context-dependency also seems to be a prototype-based category motivated by the cues languages use to construe an image-schema of discourse.

Hoover (1992), for his turn, analyzed the sentence processing strategies of 12 English and 12 Spanish-speaking subjects in parallel by computing the reading time subjects spent on each word and asking them to perform four comprehension tasks. This study is of particular importance because the stimuli were sentences in which he manipulated the number of subordinate clauses in embedded and right-branching sentences (e.g. *The clerk that the cowboy humored grabbed the guest* or *The gardener that the salesman that the carpenters that the teacher recognized greeted the clerk*). When asked

to interpret the stimuli, Spanish-speaking subjects understood embedded sentences with two subordinate clauses while English-speaking subjects did not. Both groups, however, failed to understand triply self-embedded sentences. Hoover relates his findings to Cuetos and Mitchell's (1988). Because English is a word-order-dependent language, the position of the words in the sentence determines their grammatical roles, forcing English speakers to employ a late closure strategy. Spanish speakers, on the other hand, process on-line (distributed processing strategy) since the verb and its clitics and the prepositional system carry information about the grammatical relationships among the noun phrases in general (according to Bates and MacWhinney, 1989, the accusative preposition, SV agreement, and clitic agreement, in that order, are the strongest cues to actor assignment in Spanish).

Furthermore, as with the case markings in Hungarian (MacWhinney, Bates, and Pléh, 1985), when Hoover removed all the verb inflections from the stimuli, the effect remained. That finding signals the ecological validity of the local parsing processing strategy for Spanish speakers. It also reinforces my view of context dependency and word order as a prototype-based categories with a central sub-category defined by a cluster of cognitive models.

Therefore, processing strategies—late closure or on-line—have clear effects on the kinds of sentences which speakers of English and Spanish naturally comprehend. Sentences with two or three subordinate clauses proved more difficult for the English speaker than for the Spanish speaker (see Hoover, 1992). This led Hoover to conclude that because of cognitive constraints imposed by linguistic and contextual devices, English typically employs shorter sentences and is more context dependent than Spanish (Hunt and Agnoli, 1991). In Hunt and Agnoli's words "Different languages pose different challenges

for cognition and provide differential support to cognition” (p.387). Their research on different languages, including Italian, Spanish, French, and English, shows that in highly word-order dependent languages like English, embedded structures are prone to ambiguity, posing a high cost of processing. More inflected languages like Spanish, Italian, and French, on the other hand, allow a greater number of relative clauses or right branching structures than English because on-line processing is possible.

Reading time experiments conducted by Just and Carpenter (1992) reinforce Hoover (1992), Hunt and Agnoli (1991) and other research studies that relate the amount of information linguistic systems offer to the processing strategies readers use. Working with English discourse, they show that some pieces of the interpretation puzzle only fall in place at the end of a sentence. Their study investigated the eye movement of subjects while reading sentences and indicates that the amount of time subjects spend fixating on a word is basically proportional to the amount of information provided by the word. Also, Just and Carpenter show that subjects pause at phrase boundaries to integrate information that cannot be processed until the sentence is complete. Then, they spend an amount of time at the end of sentences wrapping up meaning that they cannot construe at phrase boundaries. As Anderson said (1995), Just and Carpenter’s subjects maintained a representation of the phrases in the memory buffer because the interpretation could be wrong.

Just and Carpenter (1992) explain that result with the principle of immediacy. According to this principle, people try to extract as much meaning out of each word as it arrives, and they do not wait until the ends of sentences or even phrases to decide on how to interpret it. If there was not enough information for construing the adequate image,

they formed a less adequate one for later modification. The modification strategy implies a cost of time in terms of processing in that subjects have to modify an existing composite structure to the end of a sentence.

Therefore, reading time studies provide strong evidence for the relation between the amount of information linguistic systems offer and the processing strategy users apply. If enough information is provided, processing occurs on-line. If not, processing is either delayed to the closest boundary where integration is possible or it takes place immediately and has to be modified according to incoming information. Either late-closure strategy or subsequent reinterpretations reflect a cost in reading time.

As related to perception

Up to now, we have seen that morphological, semantic, pragmatic and cognitive factors like context dependency, processing strategies, and the amount of information a word carries affect sentence production and comprehension in terms of processing time. This sub-heading reviews studies that have related linguistic devices, perception, and the processing of information (Reinhart, 1984; Shridar, 1989; Flores d'Arcais and Schreuder, 1988). There are "organization systems that reflect properties of the human mind and restrict the way humans can process both visual and temporal or linguistic information" (Reinhart, 1984, p. 790).

Reinhart (1984) propose that the organization of narratives in temporal sequences is an extension of the principles that govern the spatial organization of the visual field into figure and ground, in other words, the principles of gestalt perception. She surveyed Labov (1972), Hopper and Thompson (1980, 1984) and other studies that investigated narrative texts and shows that the figure-ground distinction is marked linguistically at

different levels. In other words, the figure-ground distinction seems to reflect the organization and perception of language.

At the morphological level, Reinhart (1984) cites the case of the verb morphemes *-ki* and *-ka* in Swahili that distinguish the backgrounded and foregrounded textual units respectively; the Chipewyan marker *-huldu* that is attached to the whole sentence in the form of a connective, opening foreground sentences; and aspectual distinctions such as French use of the imperfect to signal backgrounded units that provide the necessary conditions to understand the sequence of events in the perfect, the foreground. At the syntactic level, she cites the use of word order in Old English. (SVO order marks backgrounded units in that language) and syntactic embedding in modern languages. As Reinhart says, material presented in subordinate clauses normally is backgrounded, providing more information concerning the main assertion.

Shifting to the organization of narratives, Reinhart (1984) demonstrates that the temporal material of the represented world provides the necessary basis for construing the foreground of the narrative text. Only events ordered on the same time axis can be presented as foregrounded. In addition, she shows that there is a striking correlation between the perceptual criteria determining the narrative foreground and those determining the figure. In her discussion of the gestalt principles of perception, she explains that what defines the figure and ground asymmetry is the functional dependency of the figure upon the ground. People perceive the same geometric figures differently depending on the angular relations and direction of the background. While the ground can determine the interpretation of the figure, the opposite is not true.

In Reinhart's (1984) view, the same dependency holds for foregrounded and backgrounded relations in narratives. The foreground is a chain of events ordered on a time axis and that chain only acquires significance because readers know the details, motivations, and circumstances that led to the chain of events. Only through the background can we understand the foregrounded events. Reinhart emphasizes that though we tend to associate the figure with the important part of the visual field or the center of attention these notions are not part of the concept of figure.

The results of the linguistic studies Reinhart (1984) surveyed reveal that we have a set of criteria, with different weights, that determine the choice of foreground. In the case of narratives the most fundamental condition is temporal order. Only textual units whose order matches the order of the events they report can serve as foreground. This condition correlates with the gestalt principle of good continuation. The principle states that we always give the highest priority to continuous contour or shape. A second criterion in the choice of foreground units is punctuality. Punctual events occupy a smaller amount of time than durative events, tending to be figures. This criterion corresponds to the gestalt principle of size and proximity. People complete the ground under the figure, and the smaller the figure, the less we have to complete. Finally, people use the criterion of closure to determine which textual elements will serve as foreground. The more closed the area is the easier it is to interpret it as a figure. Shifting to temporal relations, completed events are defined in time on both ends (beginning and end). They are closed temporally. This is a strong criterion in narrative texts because its violation will also violate criterion number one (temporal order) which is a necessary condition for

narratives. In brief, the organization of narratives is based on perceptual strategies similar to those employed in visual perception.

Reinhart (1984) reinforces McCleary's (1982) analysis of the role of transitivity in narratives (see reviews in Chapter II and III) with the tools of cognitive psychology. As McCleary says, the pragmatic function that explains transitivity may be found in a universal cognitive function: salience in the human perception of experience.

Similarly, Shridar (1989) focuses on visual perception and its correlation to linguistic descriptions. In a series of experiments, he manipulated the perceptual salience of objects—size, humanness, camera focus—to find out where more salient entities appear in subjects' descriptions of movie sequences. Results reveal that entities that enjoy great salience tend to be encoded first in discourse, across the 10 languages he examined. In highly SVO languages like English, the influence of salience in discourse led subjects to the use of object-fronted topicalized sentences.

Other studies using a range of experimental tasks have manipulated word order via perceptual salience (see reviews by Bock, 1982; MacWhinney, 1977) and previous work by Talmy (1978) and Gernsbacher and Hargreaves (1988) have demonstrated how word order is related to perception. If I say *My car is close to the bar* and *The bar is close to my car*, I profile different images because there is a change of perspective and in cognitive accessibility of the reference node (see also Chapter III about perspectivization). First mentions enjoy greater prominence because they act as reference points or 'structure building frameworks' (Gernsbacher and Hargreaves, 1988), laying the foundation of discourse.

Recent work in cognitive linguistics has also confirmed the relation between linguistic expressions and perceptual gestalt principles (Borneto, 1996; Verhagen, 1996; see details in Chapter three). Borneto claims that the shift of attention or the ground/figure asymmetry is encoded directly in the semantics of the verbs *stehen* (=to stand) and *liegen* (=to lie). While *stehen* prototypically places readers' attention on the figure (the topic or trajector), *liegen* places readers' attention on the ground (the comment or landmark). Verhagen relates word order to subjectification. Her analysis indicates that participants that comes first (first mentions) are always conceptualized independently with respect to the ones that follows. The same holds true for Gernsbacher and Hargreaves' (1988) results (see above).

Flores D'Arcais and Schreuder (1987) also found a relation between linguistic expressions and perceptual gestalt principles in experiments using lexical priming. The prime either was unrelated to the target or was related to it in terms of perceptual or functional properties. For example, the word pair *ball* and *cherry* are perceptually related because their referents have the same spherical shape. *Banana* and *cherry* are functionally related because they refer to the category fruit. Results indicate that pairs of words that share a perceptual referent are processed faster because perceptual attributes ease the access of the associated mental lexicon. The physical elements had an earlier or initially stronger activation than functional elements. Though this study treated only referential words or nouns having concrete, precise picturable referents, Flores d'Arcais and Schreuder (1987) reveal similar results for word-word pairs in a series of experiments that involve object naming and an object-object priming paradigm. In general terms, Flores d'Arcais and Schreuder's experiments showed that it is possible to be unable to recognize

a particular member of a category while at the same time being able to name the category based on the perceptual attributes of its members.

In brief, perceptual principles have proven to be rich and insightful to explain why language is organized the way it is. McCleary (1982), Reinhart (1984), Shridar (1989), Borneto (1996), Verhagen (1996), and Flores d'Arcais and Schreuder (1987) demonstrate that the way people view and choose to depict an event reflects the gestalt principles of perception.

Summary. This chapter reviewed studies on discourse processing that provided several pieces of evidence for the elaboration of this dissertation. Results from memory experiments reveal that the presence of canonical patterns speed reading times. These experiments also demonstrate that working memory or the memory that is active at a point in time to perform a specific task has a limited capacity (8 digits); its traces fade away in 20 to 30 seconds; and the faster the memory search the more items the memory can keep under focus. It is how frequently and how recently we have used a memory that determines the speed and probability of accessing it, and its level of activation. Then the chapter addressed how memory-related facts reflect in recent models of discourse processing. The models recognize the relevance of bottom-up processes and the limitations of working memory in comprehension. Then, the chapter demonstrated the role of linguistic attributes in memory and discourse processing studies. Linguistic attributes facilitate, speed, slow down or inhibit cognitive processes depending on the type of connection they establish in the network and their relation to the limitations of working memory. Anaphoric and cataphoric morphology signal continued activation of the same node in the network. Therefore, they are time-saving devices. On the other hand, lexical

connections enhance the strength of the network connections, slowing down the speed of information retrieval. This chapter also reviewed crosslinguistic studies on sentence processing. They corroborate the view that linguistic expressions are instructions to the processor as well as that the more ecologically valid cues are the more likely they will influence processing strategies in a specific language. According to crosslinguistic research, because processing strategies are responses to the typicality of linguistic cues, different linguistic attributes lead users to apply different cognitive routines. It follows that inflected languages like Spanish, Italian, French, German whose verbs and clitics carry information about the grammatical relationship allow distributed or on-line processing. Finally, reading time studies also provide strong evidence for the relation between the amount of information linguistic systems offer and the processing strategy users apply. If enough information is provided, processing occurs on-line. In brief, syntax, semantics, world knowledge, and canonical schemas conspire to guide comprehension. It is the interaction of these schemas that determine how we process language.

Some preliminary conclusions

Making this long story short, much like word order and other natural language categories, discourse seems to be an image-schema I will call COMMUNICATIVE TEXT. The schema arises from the interaction of a series of other schemas in our daily experiences with language, including a schema or text model that belongs to the conceptualizer.

The concept COMMUNICATIVE TEXT derives from research in cognitive linguistics (see Chapter III) and is soundly anchored on work on discourse processing (see Chapter IV), discourse analysis, and pragmatics (see Chapter II). Researchers such

as Lakoff (1990) have used the notion of image-schemas to explain our understanding of SPACE, MOTION, VERTICALITY, among others. Researchers such as Slobin (1982) have used it to explain the English canonical SVO word order. However, there is no research that has ever applied it to the study of discourse using the tools of Cognitive Grammar (Langacker, 1991), and combining these four areas of interest.

Cognitive Grammar has helped us to associate textual entities like nominalizations, prepositional phrases, head-modifier sequences with the trajector-landmark asymmetry, and, ultimately, with conceptualization routes. According to CG, final composite structures or abstract gestalts like the COMMUNICATIVE TEXT derive from the layering of grammatical constructions. Grammatical constructions, for their part, indicate the cognitive routes toward the image-schema.

Studies in discourse analysis led us to understand written discourse as a speech event. Similarly to speech, written discourse is interactive, goal oriented, and has a situation context. Socialization, literacy habits, and traditions relate to the shape and content of written discourse. Finally, research in discourse analysis, specifically in pragmatics, gave us the Principle of Relevance (Sperber and Wilson, 1986; 1995). This principle may help us to explain the different image-schemas discourse has in different languages. Interesting enough, there is no other study in the field that has applied the Principle of Relevance to counts of empirical data and their balance in discourse. This could have shed light on the forces that drive use of linguistic attributes in the context of a corpus.

The COMMUNICATIVE TEXT combines evidence in those areas of research. According to it, texts are interactive as much as speech and involve schemas such as

SOURCE-PATH-GOAL- DESTINATION (Lakoff, 1990). The SOURCE moves through a PATH, according to the ORIENTATION of the conceptualizer, to reach a GOAL—effective communication. To do so, it combines linguistic attributes, cognitive, pragmatic, perceptual, rhetorical, socio-linguistic factors. In reaching the GOAL, the SOURCE gets to a DESTINATION, the addressee. According to CG, each of these sub-components forms a gestalt in their own right that has other sub-components. The sub-components are activated and superimposed simultaneously to form the COMMUNICATIVE TEXT. It is the final composite structure that users rely on to recognize, process or produce any member of the category in a given language. This schema is grounded on our organization of space and on our experience with language and communication.

Furthermore, similar to other natural language categories, the COMMUNICATIVE TEXT clusters attributes of different members and exhibits prototype effects. The more inclusive member takes the center of the category and serves as a cognitive reference point. The others (non-central extensions) stand in a family resemblance relation to the established image-schema. The more the extensions meet linguistic, perceptual, and cognitive conditions in a given language, the closer they are to the central member of the COMMUNICATIVE TEXT in that language. The more they embody high-order perceptual experiences—PATH-GOAL-DESTINATION—the closer they are to the center of the category, and the faster they are recognized and processed by users.

These abstract, high-order cognitive structures also help us to understand what is universal among languages. They are the actual measures of equivalence among languages, and are what speakers of different languages share in common (see Lakoff,

1990; Shridar 1989; Reinhart, 1984). There are perceptual facts that override language differences. As Szwedek (1984) mentions in his discussion of paraphrases or the relation between active and passive constructions, *what* is said about the referent and *how* it is said differs because the presuppositions are different. However, the constructions have the same *referent* (see also Langacker, 1991; MacLaury, 1995 in Chapter IV).

Therefore, to construe a COMMUNICATIVE TEXT, language users make subjective decisions that involve the strength and ecological validity of linguistic attributes such as word order, inflections, and sentence length. They also decide how to distribute these attributes according to the allocation of attention resources, perceptual, and pragmatic factors. The more a piece of discourse resembles the text image-schema a given speech community nurtures, the more communicative and easy to process it is.

Figure 5 illustrates the image-schema I am proposing with respect to written discourse. The COMMUNICATIVE TEXT final composite structure is a complex conception or entity in which component structures such as ORIENTATION, SOURCE, PATH, GOAL, and DESTINATION are superimposed or simultaneously activated to form a single gestalt. As indicated by the heavy lines, each component has its own structure and form gestalts themselves. The dashed lines indicate interconnection of the component structures, and the arrow indicates the direction of energy flow. The heavy-line bar indicates simultaneous activation in time.

Figure 6 breaks down the components of the PATH schema. Namely, it illustrates the interaction among linguistic, socio-cultural, pragmatic, cognitive, perceptual, and rhetorical factors in the activation of a COMMUNICATIVE TEXT PATH gestalt. All the components of the composite structure and their sub-components behave as members

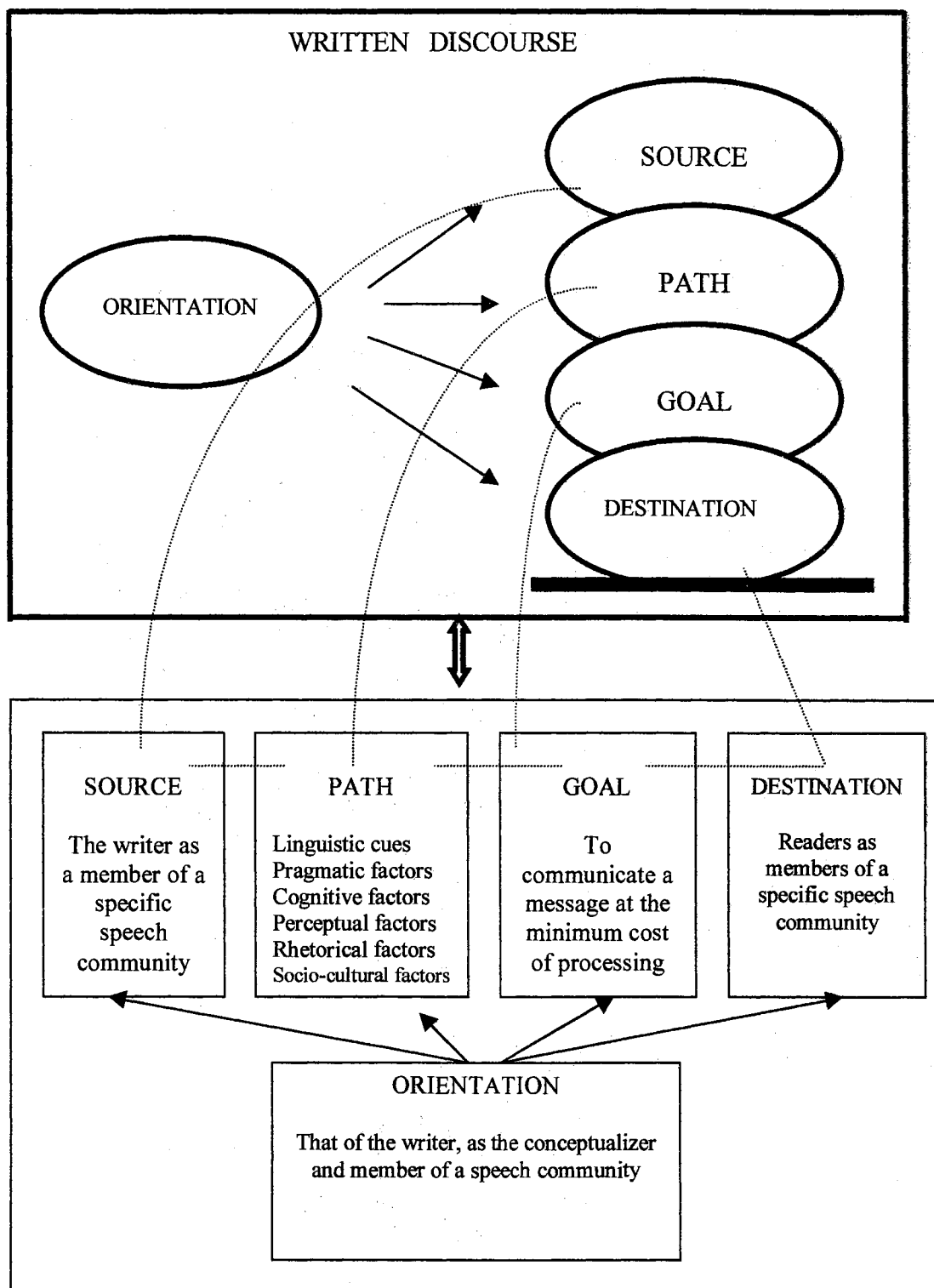


Figure 5. Texts as image-schemas: The COMMUNICATIVE TEXT.

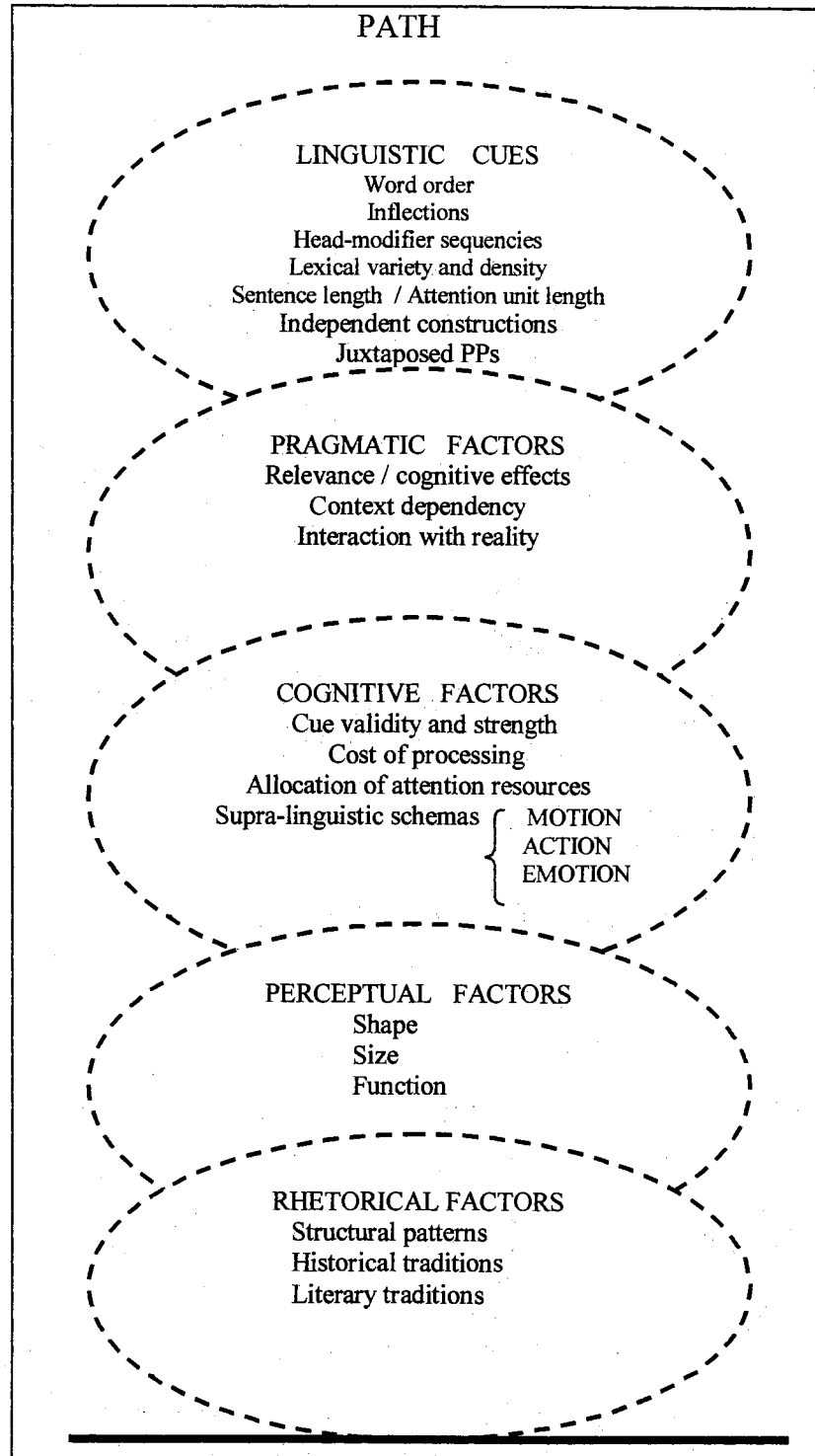


Figure 6. The PATH sub-component structure

of a family in that they all contribute to form the gestalt COMMUNICATIVE TEXT. However, the sub-components are not absolutely necessary. Their co-occurrence, though, optimizes cost of processing. This fact is supra linguistic and applies to all languages. Variants are explained by the interaction among the schemas, among the schemas and their sub-components, and among the schemas with other existing schemas, including the text model that belongs to the conceptualizer.

To conclude, each language has its own way of construing the linguistic, pragmatic, socio-cultural, and rhetorical sub-components of the PATH gestalt (see crosslinguistic studies on discourse processing in this chapter). However, language users across languages face the limited capacity of working memory when superimposing the schemas and their sub-components. Therefore, the way they combine the attributes of the PATH gestalt function as a trigger to cognitive operations such as attention activation or inhibition. Morphologic trappings, for example, keep information under the focus of attention. In doing so, they optimize processing time. Ultimately, the COMMUNICATIVE TEXT yields maximum cognitive effects.

Chapter VII applies the theoretical notions developed here to the empirical analysis of Brazilian Portuguese and English discourses. Before I do that, however, I will present some facts related to Brazilian Portuguese to ground both the empirical study of BP texts and the analysis that follows thereafter.

CHAPTER V
FACTS ABOUT BRAZILIAN PORTUGUESE

Overview

In this Chapter I briefly describe some facts about Brazilian Portuguese (BP), contrasting them to languages that have been investigated in studies on sentence processing (see Chapter IV) and that belong to the Romance family such as French, Spanish, and Italian. Particularly, I discuss word order, agreement, and head-modifier relations in BP. In addition, I draw on my literature review to make theoretical hypotheses about possible attributes of a BP COMMUNICATIVE TEXT.

Word order in BP

Traditional typologies consider Brazilian Portuguese (BP) a free order language (Shepherd, 1987; Camara, 1977; Cunha and Cintra, 1985). Indeed, because morphology in BP expresses many categories, Portuguese word order is more flexible compared to English. Actually, the order of elements in spoken and written BP becomes grammatically irrelevant when in competition with aesthetics. For example, all of the variations below could occur in Portuguese (see Cunha, 1987; Dantas, 1983; my own observation as a native speaker):

SVO	<i>João</i>	<i>deu</i>	<i>o</i>	<i>livro</i>	<i>... a</i>	<i>Maria.</i>
	John	gave (3rd.p.s.)	the (m.s.)	book (m.s.)	to	Mary.

VSO	<i>Deu</i>	<i>João</i>	<i>o</i>	<i>livro</i>	<i>.. a</i>	<i>Maria.</i>
	Gave (3rd.p.s.)	John	the (m.s.)	book (m.s.)	to	Mary.
OSVO	<i>A</i>	<i>Maria, João</i>	<i>deu</i>	<i>o</i>	<i>livro.</i>	
	To	Mary John	gave (3rd.s.)	the (m.s.)	book (m.s.)	
OSVO	<i>O</i>	<i>.. livro,</i>	<i>João</i>	<i>deu</i>	<i>a</i>	<i>Maria.</i>
	The (m.s.)	book (m.s.)	John	..gave (3rd.p.s.)	to ..	Mary.
V(s) O	<i>Deu</i>	<i>o</i>	<i>livro</i>	<i>a</i>	<i>Maria.</i>	
	Gave (3rd.p.s.)	the (m.s.)	book (m.s.)	to	Mary.	

However, the free word order of BP does not imply indifference to semantic value or to the thematic structure of discourse. Though the SVO order occurs with great frequency and enjoys high ecological validity (see Dantas study of word order in Portuguese, 1983), BP speakers consider all the other order alternatives equally natural if compelled by the discourse context, prosody, or the thematic structure. The same holds true for its sister languages, Italian and Spanish, and to a lesser extent, to French. Therefore, word order in Portuguese is simply **more flexible** compared to English.

That is so because, contrary to English and similar to Spanish, Italian, and French, BP speakers may generally rely on verbal and nominal inflections to identify the subject and disambiguate sentences. Compare the examples below:

SVO	<i>O</i>	<i>... mar</i>	<i>engoliu</i>	<i>as</i>	<i>rochas.</i>
	The (m.s.)	sea (m.s.)	engulfed (3rd.p.s.)	the (f.s.)	rocks (f.pl.)
SVO	The sea engulfed the rocks.				

OVS *O mar, engoliram as rochas.*
 The (m.s.) sea (m.s.) engulfed (3rd.p.pl.) the (f.pl.) rocks (f.pl.)
 SVO **The rocks engulfed the sea.**

While in English the agent necessarily appears in first position, in BP, verb morphology signals who is doing what to whom and keeps the referent active within the episode (see my review of Lambrecht's study about French spoken discourse—1987—in Chapter IV). It doesn't really matter where the expression is located syntactically. In spoken language, pragmatics and prosody would corroborate such an interpretation (see Dantas, 1983, on the effects of prosody). Like in French, full NPs or stressed pronouns in subject position are rare (my own observation). Verb-morphology keeps that information active after the first-mention and optimizes processing time.

Nevertheless, if it were not for the singular-plural distinction between the subject and the object reflected by verb morphology, BP would have had to rely on word order as well:

O mar engoliu a rocha.
 The (m.s.) sea (m.s.) engulfed (3rd.p.s.) the (f.s.) rock (f.s.)

A rocha engoliu o mar.
 The (f.s.) rock (f.s.) engulfed (3rd. p.s.) the (m.s.) sea (m.s.)

Both *rocha* and *mar* can act as agents or sources of energy in the action chain. A sentence like—*Wellington beat Napoleon*—(proposed by Clark and Clark, 1977) is

another example. In this sentence both names could act as agents if it were only for morphological cues. Either *Wellington* or *Napoleon* represents a third-person-singular animate entity that could equally function as the source of energy in the action chain metaphor. The relational predicate is attached to a past-third-person-singular ending *-eu* that could apply to either source of energy. It is word order that assigns the name that appears first the agent role (see my review of Shridar, Chapter IV, on the role of perception in narrative re-telling). Observe:

Wellington	beat	Napoleon.
<i>Wellington (3rd.p.s)</i>	<i>venceu (3rd. p.s.)</i>	<i>Napoleão (3rd. p. s.)</i>
S	V	O
Napoleon	beat	Wellington.
<i>Napoleão (3rd.p.s.)</i>	<i>venceu (3rd.p.s.)</i>	<i>Wellington (3rd. p. s.)</i>
S	V	O

Therefore, word order in written BP is flexible to the degree that morphological trappings disambiguate reference. In the cases that morphology disambiguates reference, it keeps information active throughout stretches of discourse by establishing relations between arguments and predicates. In other words, morphology cues meaning and frees language users of some processing constraints. In the next subheading, some grammatical markings are analyzed under the focus of that assumption.

Grammatical markings

One of the most striking differences between Portuguese and English is that Portuguese is a syntactic language with an extremely rich morphology while English is an analytic and morphologically poor language. In Portuguese, each element may have

several functions, whereas in English each element carries one function. For example, in the word *gostaríamos* (= *we would like*), we have: *Gost* (=like) *a* (=theme of the first conjugation) *ri* (=TEN; preterite future-ASP; imperfective-MOOD; indicative) *amos* (=PERSON; 3rd.-NR; plural). Because verb morphology potentially disambiguates cues, both written and spoken discourse commonly delete the subject. Utterances like (S) *Chove* (=It is raining) or (S) *Chegou* (=he/she has arrived) occur frequently. Besides, there are those sentences grammarians call “clauses without subject” (=null subject) which correspond to existential constructions in English and for which Portuguese requires no such thing as a “dummy” subject like *it* or *there*:

(S) VOLOC Há uma moça à porta (There) is a young girl at the door.

Portuguese is then what linguists call a pro-drop language much like Italian and Spanish and unlike English and French, which require overt subjects, including dummy subjects like *it* in *it is raining* (= *Il pleut* in French).

As to subject-verb agreement, while English marks only third-person singular (*drinks*), requiring subject-pronouns to disambiguate reference, Portuguese inflects each person-number form. Ambiguity occurs in the third person singular or plural in relation to gender and reference (e.g. *canta* = *he/she sings* or *you sing* for the colloquial form of address *você*). It also occurs with the first and third person singular of verbs of the second and third conjugation. In those instances, word order (see my discussion above) and discourse features such as context become disambiguating cues.

Additionally, Portuguese always attaches gender and number markers to adjectives, nouns, and pronouns (it lacks the neuter pronoun *it*) while English lacks gender markers and does not inflect adjectives either for number or gender and it carries a neuter-pronominal form.

The grammatical markings described in this brief description of BP conspire to pack information into single forms and to establish links among entities that are figure and those that are ground. Those links are like anaphoras, keeping information active in readers' minds. In addition, the markings also make the entities that they elaborate more salient.

Next, this Chapter discusses head-modifier relations, another attribute that may contribute to the image-schema of a Brazilian Portuguese COMMUNICATIVE TEXT.

Head-modifier relations

BP users may also rely on head-modifier relations when processing information. The head first cue corresponds to the sequence of events in perception, a variable that Shridar (1989) has shown to affect sentence production. Brazilian Portuguese is a head-initial language. In a phrase like *cadeira de balanço* (=rocking chair), a BP speaker immediately assigns a role to the word *cadeira* (=chair) upon hearing it, adopting a local processing strategy. English, on the other hand, is a head-final language. The English speaker has to wait until the head or the profile determinant entity (the trajector = TR) is enunciated to construe the image (see my review of Cognitive Grammar in Chapter III). Another way of putting it is that, in Portuguese, what is instantiated first is a more central member of the category chair—*cadeira*—which is the region or domain that is perceptually more salient or autonomous. Once this autonomous entity is primed, it is available for any modification by a dependent entity (the landmark = LM), which profiles a relation between an abstract

domain and the autonomous entity (the more central member). In this case, the dependent entity is a PP (=de *balanço*), an entity that carries the e-site or the profile determinant structure (= *balanço*). I suggest then that the perceptual-motor constraints in constructions that involve head-modifier relations (including possessives) are weaker in Portuguese than in English, favoring an on-line processing strategy.

The examples below further reinforce my point of view. While in BP the autonomous entity appears first in all instances, in English, the dependent entity appears first:

TR LM LM	O departamento Brasileiro de mineração	LM LM TR	Brazil's mineral department
TR LM	O carro de Maria	LM TR	Maria's car
TR LM LM	A situação política do Brazil	LM LM TR	Brazilian political situation

However, a cross-examination of crosslinguistic research on sentence processing (Bates and MacWhinney, 1989; Hoover, 1989; Lambrecht, 1987; see other studies in Chapter IV) indicates that Portuguese lacks some markers that in other Romance languages contribute to on-line processing. For example, unlike Spanish, BP does not have the preposition *a* as a marker of human direct object. This suggests that Portuguese is “less” of a free word order language than Spanish, lying somewhere in between English and Spanish in the word-order-dependency category continuum. Further research may fine grain our understanding of attributes that differentiate Romance languages in terms of word order. Figure 7 tentatively illustrates the continuum. In it, Italian appears in a position similar to Portuguese, and French in a position closer to English:

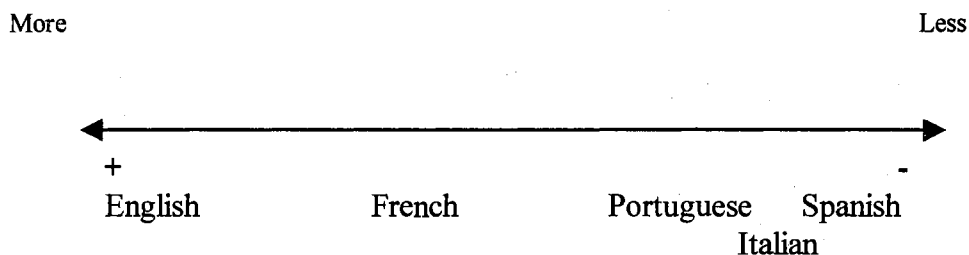


Figure 7. Romance languages: Word-order-dependency continuum.

Furthermore, like in Spanish, French, and Italian, in Brazilian Portuguese the clitic *se* also attaches to certain verbs assigning an agent, patient or both roles to the subject and disambiguating reference. For example, upon hearing *João se* ... (=John caused himself to...), BP users know that John is both the source and the absorber of energy in the action chain. It is possible to assign roles immediately because of the clitic *se*. In English, however, on-line assignment of roles would not be possible. Even in sentences like *John sank...* (example adapted from Hoover, 1992), English users would need to wait until the end of the sentence to assign roles or to know if John went under the water or whether John caused something to sink. If we change the orientation of the sentence, having the inanimate thing as the source of energy (=The boat sank), there is no way, in English, to know if the boat sank by itself, being the exclusive source of energy, or if there was an animate, prototypical source of energy in the action chain that caused it to sink. That does not happen in Portuguese. Like in the other Romance languages, BP can mark the verb with the clitic *se* to eliminate potential ambiguities, favor on-line processing, and mark an inanimate thing as the source of energy (=The boat sank by itself):

[John caused the boat to sink]	[The boat sank by itself]
John <i>sank</i> the boat	The boat <i>sank</i> .
João <i>afundou</i> o barco	O barco <i>se afundou</i>
Juan <i>hundió</i> el barco	El barco <i>se hundió</i>

Example from Hoover (1992, p. 278; my translation to Portuguese)

If Portuguese is close to Italian and Spanish descriptively and if speakers may rely most of the time on clitics, verbal and nominal inflections as well as on cognitive and perceptual variables to assign semantic roles on-line, BP COMMUNICATIVE TEXTS will typically use lengthier sentences than English COMMUNICATIVE TEXTS. Such a hypothesis is grounded on previous research on sentence processing in languages descriptively similar to Portuguese (namely, Spanish and Italian; see Chapter IV; Hunt and Agnoli, 1991 in special), in existing studies in BP (Oliveira 1997a; Oliveira 1997b), and on my own argument for a cognitive rationale to explain written discourse using the tools of CG. Sentence length seems to be a function of linguistic, cognitive, perceptive, and pragmatic factors.

CHAPTER VI

THIS STUDY

Overview

This chapter describes the rationale and facts about the study, comments on the COMMUNICATIVE TEXT image-schema, and details the method used to investigate Portuguese and English discourse.

Rationale

The review of the literature (Chapters II, III, and IV) has shown how studies have focused on discursive features like cohesive ties, frequency of grammatical categories in a specific genre, patterns of information structure, and topic continuity. It has also shown how social, cultural, pragmatic, and cognitive factors relate to differences across languages and to an image-schema that members of a speech community share and use as a point of reference toward comprehension.

However, most studies in discourse analysis have treated English texts only. Others that have dealt with crosslinguistic research have addressed languages like Italian, French, Spanish, German, Chinese, and Japanese, contributing little to our understanding of Brazilian Portuguese as it relates to the English language and to cognitive processes. Finally, the review of the literature focused on a new paradigm, cognitive linguistics, and on the descriptive tools of Cognitive Grammar. Both seem to offer a consistent and adequate apparatus for analysis of discourse and crosslinguistic research. Nevertheless,

the review of the literature also fails to show studies in cognitive linguistics that have explained prototypicality with reference to functional considerations and other features typical of studies in the field of discourse analysis using the mechanisms of CG. As Langacker (1991) envisaged, to obtain explanatory adequacy, linguistic phenomena should be examined from psychological, discourse, sociolinguistics, and ecological perspectives.

This dissertation provides such a contribution. It combines discourse analysis techniques with Cognitive Grammar and compares discourse in Brazilian Portuguese to English discourse. Specially, it has been conducted with four goals in mind. First, I want to analyze institutional expository Portuguese and English texts to find the linguistic attributes which are common (or highly frequent) in this genre. I characterized the genre as “institutional” because its major goal is to sell a positive image of the oil companies that publish them. The investigation involves proposing an image-schema for what I have called a COMMUNICATIVE TEXT category in each language. Second and more importantly, I want to demonstrate that these schemas emerge from users’ need to balance cost of processing. To do so, I will draw on cognitive linguistics, cognitive psychology, and discourse analysis frameworks as described in my review of the literature. Third, I would like to show that this type of analysis provides a sound framework to explain discourse translation, comparing findings of texts written in English to those of texts translated to English, as well as to the original source in BP. Fourth, I shall draw conclusions from my research to propose a theory of discourse analysis that combines competence with performance, psychologically explaining rhetorical differences among languages.

Facts about the study

Following Blass (1990), Sperber and Wilson (1986; 1995), Hunt and Agnoli (1991), Chafe (1994), Biber (1988; 1994), Grabe (1987), Lakoff (1990), Langacker (1991), Bates and McWhinney (1989), and Shridar (1989), I am interested in observing the regularities and frequencies of linguistic phenomena in English and Portuguese and in understanding how these outward signs represent a pattern of events including a mental image.

The theoretical inferences I will make will be based on observed behavior (published written discourse) or manifestations of textual attributes. This type of data may only be interpreted as indirect evidence for the ecological validity of the COMMUNICATIVE TEXT category in each language. However, I will ground those inferences on statistical findings and on existing research on sentence processing on languages historically and descriptively related to Portuguese like Italian, Spanish, and French (see Chapter III).

The statistical analysis, though, will simply provide the empirical evidence for mechanisms that go on in the mind and that constrain language use, my main concern. Contrary to Brown and Yule (1983), Biber (1988; 1995) and others that have focused exclusively on empirical evidence and on a functional analysis, and following Chafe (1994) and Blass (1990), I see the most interesting issues in discourse analysis as WHY a text is structured in a particular way. A qualitative analysis of linguistic attributes in the light of cognitive linguistics will provide us this information. Again, textual facts reflect pragmatic and psychological considerations.

The COMMUNICATIVE TEXT image-schema

In spite of all the research that has been conducted in discourse analysis, contrastive rhetoric, social and cultural differences, claims hinging on the weak version of the

Whorfian hypothesis and language specific textual structures, the field still needs an adequate explanation about what a text is and how to accommodate textual differences within and across languages.

Like any other linguistic expression, a text has entities—topic, grammatical constructions, participants, outcomes—that can be associated with different aspects of the trajector-landmark or the figure-ground asymmetry. Each entity can be assessed against the total gestalt (I direct the reader to Chapter III, Figures 5 and 6). Within that gestalt, there is a prototype that may coincide or not with the whole and that serves as a reference point to comprehension of other manifestations of the text. As I have mentioned in Chapter III, this image-schema is what I call COMMUNICATIVE TEXT. In other words, an abstraction of traits shared by a great number of texts in a language that are mapped against the ORIENTATION-SOURCE-PATH-GOAL-DESTINATION schemas.

Members of a given speech community have created the COMMUNICATIVE TEXT, conventionalized it by usage, and have been using it as a reference point toward the construction of other manifestations. Each new text one reads is nothing more than a metaphorical expression of previous instantiations or extensions of the schema that speakers share. Different texts within the schema manifest the SOURCE-PATH-GOAL-DESTINATION schemas (Lakoff, 1990) according to their interaction with the conceptualizer's ORIENTATION. As such, texts represent the structure of a conceptualizer's everyday bodily experiences. That is, they represent how a conceptualizer moves from one point in space to another, and in this motion, s/he follows the shortest PATH possible to reach a DESTINATION in the shortest time possible. Therefore, texts that fit the COMMUNICATIVE TEXT image-schema free language

users' working memory of unjustified processing constraints. They do not exist per se, but consist of events, processes, experiences located in time and space in a given psychological reality (Lee, 1996).

Now, each language community emphasizes or suppresses different attributes of the COMMUNICATIVE TEXT schema, depending on external motivation and on cognitive variables brought to bear by grammatical constructions typical of that language. This understanding explains why the COMMUNICATIVE TEXT varies from language to language. Van Dijk (1988), followed by Duszak (1995), were the first ones to propose similar views in relation to editorials (see Chapter IV).

In this dissertation I take those pioneering studies one step further, by combining principles of cognitive linguistics and the framework of CG with studies on information flow and discourse processing. I propose the COMMUNICATIVE TEXT image-schema, one of those schematic structures proposed by Lakoff (1990) that embody various metaphors—an image-schema. I emphasize that an image-schema has a sound cognitive anchor and though different combinations of its attributes may take place, the composite structure remains constant.

It is true that language users may not consciously know how to combine linguistic attributes into composite schematic-structures. However, as Delbecque (1996) indicates, they rely on a mental replay of the image that allows them to use the structures intuitively. Langacker (1991) insistently points out that a structure only sanctions itself to the degree it is solidly established and therefore salient in the language. The more salient or frequently used it is, the more energetically charged it will be in the mind of language users.

Method

To investigate discourse in Portuguese, English, and discourse translation, I analyzed 20 institutional expository texts in Portuguese and English translation, and 20 in English, totaling 6371 words in Portuguese, 6022 in discourse translation, and 6398 in English. My analysis focused particularly on the cognitive implications brought to bear by linguistic devices such as nominalizations, location and time adverbials, prepositional phrases in sequence, content words, sentence length, attention unit length, and number of independent clauses. Also, I observed the presence of word-order effects, namely head-modifier and head-complement relations, inflections, SV agreement, and structures that favor embedding and right branching structures. The co-occurrence of these events across the research conditions may be related to processing load.

Permission to reprint the texts in full was kindly given by the five different publishers, Petrobras, Exxon, Shell, Texaco, and Phillips. Below I describe the data, procedures, and type of analysis I employed to provide a schematic notion of discourse in the three research conditions.

Data

Table 1 displays the profile of the data. It consists of 20 institutional expository texts in Brazilian Portuguese from the Public Affairs Service of Petrobras, the state owned oil company in Brazil and 20 institutional expository texts in English from the Public Affairs Service of Exxon, Texaco, Phillips Petroleum, and Shell, multinational oil companies operating worldwide (a sample of the texts appear in Appendix B). The texts

TABLE 1
DATA PROFILE

TEXTS n = 60			
COMMON FEATURES			
GENRE			Institutional expository
MAIN DISCOURSE ENTITIES			Oil companies (the publishers)
TOPIC			Producing oil and contributing to a sound environment and community life
READERSHIP			Oil industry businessmen / companies' employees and shareholders/ general public
GOAL			To sell a positive image of the company
ORIENTATION			That of the publishers, the oil companies

RESEARCH CONDITIONS	TEXTS	NUMBER OF WORDS	RANGE OF WORDS
BRAZILIAN PORTUGUESE	20	6371	74-867
ENGLISH TRANSLATION	20	6022	60-852
ENGLISH	20	6398	68-831

TABLE 1 (Continued)

DATA PROFILE

ENGLISH					
n = 20					
SOURCE	PUBLICATION	YEAR	SELECTED TEXTS		# WORDS
Shell	<i>The Test of Tomorrow</i>	1993	1. Shell initiative	(p.3)	520
			2. Existing operations	(p.6)	210
			3. Past activities	(p.8)	186
	<i>Shell and the Environment</i>	1995	4. Environmentally related expenditures	(p.4)	162
			5. Marine pollution	(p.9)	262
			6. Case study—Improving techniques in the upstream	(p.9)	269
			7. Forestry	(p.12)	179
			8. Case study—Closing the loop—recycling plastic products	(p.13)	162
			9. Case study—Environment conservation	(p.17)	416
Exxon	<i>Exxon and the Environment</i>	1975	10. Blowout school	(p.40)	476
			11. Return of the terns	(p.106)	355
			12. To prevent a spill go out and look	(p.69)	625
	<i>Exxon 1990 Annual Report</i>	1990	13. Alaska update	(pp. 26-27)	537
			14. Environment safety overview	(pp. 28-29)	497
Phillips	<i>1992 Annual Report</i> <i>Good Citizenship: A Key to Top Performance</i>	1992	15. Safety	(p.25)	97
			16. Company takes new management approaches	(p.27)	157
			17. Company helps shape public policy	(p.27)	68
Texaco	<i>1979 Annual Report</i>	1979	18. Teams give employees more authority	(p.27)	216
			19. Texaco's 40 th . year with the Met	(p.29)	173
			20. Special report: Alternate Energy	(p.30)	831

TABLE 1 (Continued)

DATA PROFILE

BRAZILIAN PORTUGUESE (BP)				
n = 20				
SOURCE	PUBLICATION	YEAR	SELECTED TEXTS	# WORDS
Petrobras	<i>Conheça a Produção de Petróleo na Amazônia*</i>	1992	1. Conheça a produção de óleo na Amazônia	147
			2. Uma longa história	624
			3. A aventura da exploração	427
			4. A perfuração e seus desafios	420
			5. Produção e escoamento	629
			6. A base de Urucu	441
			7. A preocupação com o meio ambiente	867
Petrobras	<i>Meio Ambiente, Qualidade e Segurança*</i>	1992	8. Petróleo, um velho conhecido	138
			9. Esforço para preservar o futuro	158
			10. Na exploração, a convivência harmônica com a natureza	140
			11. Perfurações: Locações cuidadosamente planejadas	418
			12. Produção: Inovações tecnológicas	221
			13. No refino, a qualidade cada vez mais apurada	220
			14. Transportando com segurança	256
			15. Engenharia e meio ambiente	74
			16. Distribuição de produtos com qualidade e segurança	150
			17. Treinamento e intercâmbio com a comunidade	94
			18. Incentivo à vida natural	228
Petrobras	<i>Relatório Anual 1995</i>	1995	19. Recursos Humanos (p. 51)	436
			20. Relações Institucionais (p. 55)	283

* Pagination was not available.

TABLE 1 (Continued)

DATA PROFILE

ENGLISH TRANSLATION				
(n=20)				
SOURCE	PUBLICATION	YEAR	SELECTED TEXTS	# WORDS
Petrobras	<i>Oil Production in the Amazon*</i>	1992	1. Oil production in the Amazon	138
			2. A long history	624
			3. Oil exploration—an adventure	359
			4. Drilling and its challenges	391
			5. Production and outflow	608
			6. The Urucu base	412
			7. Concern with the environment	852
Petrobras	<i>Environment, Quality, and Safety*</i>	1992	8. Oil, a friend from way back	140
			9. Striving to safeguard our future	180
			10. Exploration in tune with nature	178
			11. Careful planning of drilling sites	365
			12. Technological breakthroughs in production	158
			13. Refining: Enhanced quality, a permanent goal	195
			14. Transporting safely	197
			15. Engineering and the environment	60
			16. Secure delivery of top-rate oil products	159
			17. Education and community exchange	84
			18. Lending nature a hand	251
Petrobras	<i>Annual Report 1995</i>	1995	19. Human resources..... (p. 39)	401
			20. Public affairs.....(p. 44)	270

* Pagination was not available.

in the English translation research condition (n=20) are the same selected for the Portuguese research condition which were translated to English and published by Petrobras.

Selection criteria. To select the texts within those publications, I matched them according to text type (exposition), genre (reportages and annual reports), sub-genre (I labeled it “institutional” discourse because of the GOAL and ORIENTATION of discourse), number of words, topic, sources, and audience. It follows that all the selected texts are institutional expository texts. That is, their GOAL is to inform the readership—oil industry businessmen, the companies’ employees and shareholders, and the general public—about the oil companies’ activities. They also aim at persuading readers that in producing oil and oil products, the companies put their best effort to contribute to a sound environment and community life. Therefore, the ORIENTATION of all selected texts is that of the oil companies or the publishers themselves. In addition, the oil companies appear as the main discourse entities in all selected texts and the number of words ranges from 60 to 867 words. Actually, I paired texts in Brazilian Portuguese and English according to their number of words. That is, if I selected a text in BP that had 420 words, I would pick up one in English with an approximately similar number of words, for example, 416. Also, the articles or texts I selected have a beginning, middle, and end. The selection criteria derived from findings of Biber (1988; 1995), Grabe (1987), and from other evidence described in Chapter II. The two next sub-headings provide full details of the corpus.

Brazilian Portuguese and English translation data

The Brazilian Portuguese and English translation data were published by the Public Affairs Service of Petrobras and are part of institutional brochures frequently issued by that department to promote the company both to the Brazilian (BP discourse) and to the international public (English translation discourse). Their main function is to sell a positive image of the company and its services, having for readership, businessmen of the oil industry, shareholders, employees of the company ranging from engineers to clerks and the public in general. Therefore, *Petrobras* is the main topic-entity in discourse. Given the broad readership of the texts, they should be as close as possible to the COMMUNICATIVE TEXT category in BP and English translation to fulfill their function.

According to the selection criteria, I gathered three brochures from Petrobrás: (1) *Conheça o petróleo na Amazônia* (=Oil Production in the Amazon, 1992), (2) *Meio ambiente, qualidade e segurança* (=Environment, Quality, and Safety, 1992), and (3) *Relatório anual 1995* (= Annual Report 1995). These texts were originally written in BP and later translated to English, covering Petrobras drilling operations in the Amazon and the company's concern with environmentally safe operations and quality of services. From brochure # 1, I selected 11 texts (n=11) texts. From brochure # 2, seven texts (n=7); from the annual report two (n=2). The total number of words of texts in BP is 6371. In English translation those same texts were used, totaling 6022 words.

English Data

The English data came from four equivalent sources: the Public Affairs Department of Exxon Corporation; Phillips Petroleum; Texaco Inc.; and Group Public Affairs, SIPC,

Shell Centre. The three first companies are American and the last English-Dutch. All of them are involved in exploration and production of oil, gas, coal, and other chemical businesses overseas. Though translation to other languages is available from the sources, in the present study I analyzed only the originals, that is, the English texts, totaling 20 texts and 6368 words. Following the selection criteria, the texts cover aspects of oil production and environmental protection in different sites where those companies operate, and the companies function as the main discourse entities.

From the Public Affairs Department of Exxon Corporation, I collected a magazine—*Exxon and the Environment* (1975)—and an annual report—*Exxon 1990 Annual Report*. The magazine is a collection of articles published by Exxon Corporation and its affiliates over the last several years to show the public that environmental protection has become an integral factor in Exxon's operations. Among them, I selected three (n=3): "Blowout School" (p.40) with 476 words; "Return of the Terns" (p.106) with 355 words; and "To Prevent a Spill go out and Look" (p.69) with 625 words. From the *Annual Report*, I chose two articles (n=2) that bring up the environmental topic: "Alaska Update: 1990" (pp. 26-27) with 537 words and "Environment Safety Overview" (pp. 28-29) with 497 words.

From Phillips Petroleum Special Projects Department, I selected sub-headings from an article published in the *1992 Annual Report*: "Good Citizenship: A key to top performance" (*Phillips Petroleum Co. 1992 Annual Report*, pp. 25-27). Namely, I analyzed the following sub-headings: (1) "Safety" with 97 words; (2) "Company Management Approaches" with 157 words; (3) "Company Helps Shape Public Policy" with 68 words; and (4) "Teams Give Employees more Authority" with 216 words.

According to the selection criteria the texts highlight Phillips' management approach and contributions to a better environment and community life.

From *Texaco's 1979 Annual Report* issued by Texaco Inc., I selected two articles (n=2) that, like the other sources, inform shareholders about the company's activities and social programs. They are "Texaco's 40th year with the Met" (173 words) and "Special Report: Alternate Energy" (831 words).

Finally, I gathered six texts from (1) *Shell and the environment* (1995) and three from (2) *The test of tomorrow* (1993) issued by Shell Centre. From brochure # 1, I analyzed "Environmentally Related Expenditure" (p.4, 162 words); "Marine Pollution" (p.9, 262 words); "Case study—Improving Techniques in the Upstream" (p.9, 269 words); "Forestry" (p.12, 179 words); "Case study—Closing the Loop—Recycling Plastics Products" (p.13, 162 words); and "Case Study—Environmental Conservation" (p.17, 416 words). From brochure # 2, I analyzed "Shell Initiatives" (p.3- 4, 520 words); "Existing Operations" (p.6, 210 words); and "Past Activities" (p.8, 186 words).

I reinforce that my rationale for choosing the cited articles within each publication rests on six reasons. (1) In all of them, the main function of discourse is to promote the business of Petrobras, Exxon, Phillips, Texaco, and Shell. (2) They all aim at a broad readership. Though the *Annual Reports* have been primarily designed to inform shareholders about the respective companies' performances during the indicated fiscal years, the articles I chose within them could well appear in a magazine directed to a general audience interested in oil companies' operations or in how they have been helping communities and contributing to a sound environment. (3) They all have their respective publishers as the main discourse entities. (4) They all share similar themes, language, and

genre (institutional expository discourse). (5) Each text in English pairs with a text in BP in terms of number of words. (6) They all have an introduction, a middle, and a conclusion. By using texts that come from similar sources (petroleum companies), have the same GOAL, ORIENTATION, genre, and topics, I improve the validity and reliability of this study's results. I also reinforce that I balanced the number of texts and words in each language (Brazilian Portuguese and English). Any differences in raw frequency counts caused by length were neutralized by normalizing the frequencies (I further explain this procedure under the Data Analysis sub-heading).

Procedures

To get at an image schema of written expository discourse in BP and English, I drew on the theoretical framework outlined in the introductory chapters (Chapter I and II) and further detailed in my discussion of the cognitive paradigm (Chapter III) and discourse processing (Chapter IV). Namely, I selected eight lexical and syntactic events in the data according to their production and comprehension demands in the communicative event (their processing functions), examined their distribution as well as how that distribution construed text image-schemas in English and BP written discourse. That research strategy allowed me to test the applicability of Chafe and Danielewicz's (1987) and Biber's (1988; 1995) findings to those of this dissertation. In that I do so, I add to our knowledge of written language properties in English, explore them in Brazilian Portuguese and English translation, and provide a cognitive crosslinguistic account of the role of linguistic events in written discourse.

Data analysis

Data were analyzed both quantitatively and qualitatively for greater detail and depth. Emulating Chafe and Danielewicz (1987), I first manually counted the occurrences of nominalizations, time and location adverbial phrases, number of running words in attention units and sentences, number of content words, sequences of two or more prepositional phrases, and number of independent clauses. Second, I computed the type-token ratio. Third, I ran descriptive statistics such as means, medians, standard deviations (SD's), and ranges for the frequencies of the selected linguistic attributes in the data. That procedure allowed me to examine the overall distribution of the selected events in each language data set. Descriptive statistics also allowed me to roughly estimate the level of syntactic complexity by providing the median number of words in a sentence. Given that the number of words generally relates positively to the number of dependent clauses, sentence length provides a rough estimate of syntactic complexity. Fourth, I compared the numeric differences in Brazilian Portuguese and English and checked if they were statistically significant by means of a two-tailed Mann Whitney U-test for independent groups. The U-test compares the sums of ranks of each linguistic attribute in the two groups. Results were considered significant at the 0.05 level. Fifth, I compared the three research conditions—English, Brazilian Portuguese, and English translation. To do so, I ran a Friedman two-way analysis of variance for dependent groups on the rank sums of each linguistic attribute in the three groups. A test for dependent groups was computed because the Brazilian Portuguese and English translation data are repeated measures. That is, they convey the same conceptual structures in different codes. Again, results were considered significant at the 0.05 level. In addition, the post-hoc Nemenyi test

located where the differences occurred and provided the Kendall Coefficient of Concordance among the groups.

Given that the texts vary in length, all counts taken from the texts were normalized to 1000 words, except for the type-token ratio. To obtain a normalized count, the actual frequency count was divided by the total number of words in the text and then multiplied by 1000. For example, the normalized frequency of nominalizations in text 1 in BP is:

$$(6 \div 140) \times 1000 = 42.8$$

That procedure avoided any bias and allowed comparisons across texts and with previous studies. Also, numbers were all rounded to the second decimal place.

Finally, using the notational devices and principles of cognitive grammar as advanced by Langacker (1991), I qualitatively examined how word-order effects, namely head-modifier and head-complement relations, inflections, SV agreement, syntactic complexity and the quantitative results cited above related to effort of processing in the collected data. In the next sub-heading, I specify how I coded each of the linguistic events. The procedures and analyses used in this study are outlined in Figure 8.

Coding of the linguistic events

By surveying existing studies (Chafe, 1982; Chafe and Danielewicz (1987), Dantas-Whitney and Grabe (1989), Biber (1988; 1995), Hunt and Agnoli (1991) and using my own knowledge about written discourse in English and Brazilian Portuguese (I am a native speaker of Brazilian Portuguese), I identified linguistic attributes that could potentially shed light on our understanding of how attributes account for text schemas in BP and English. Then, following my review of the literature on discourse processing (see Chapter IV), I selected eight events based on the effort they demand in production and

PROCEDURES AND ANALYSIS

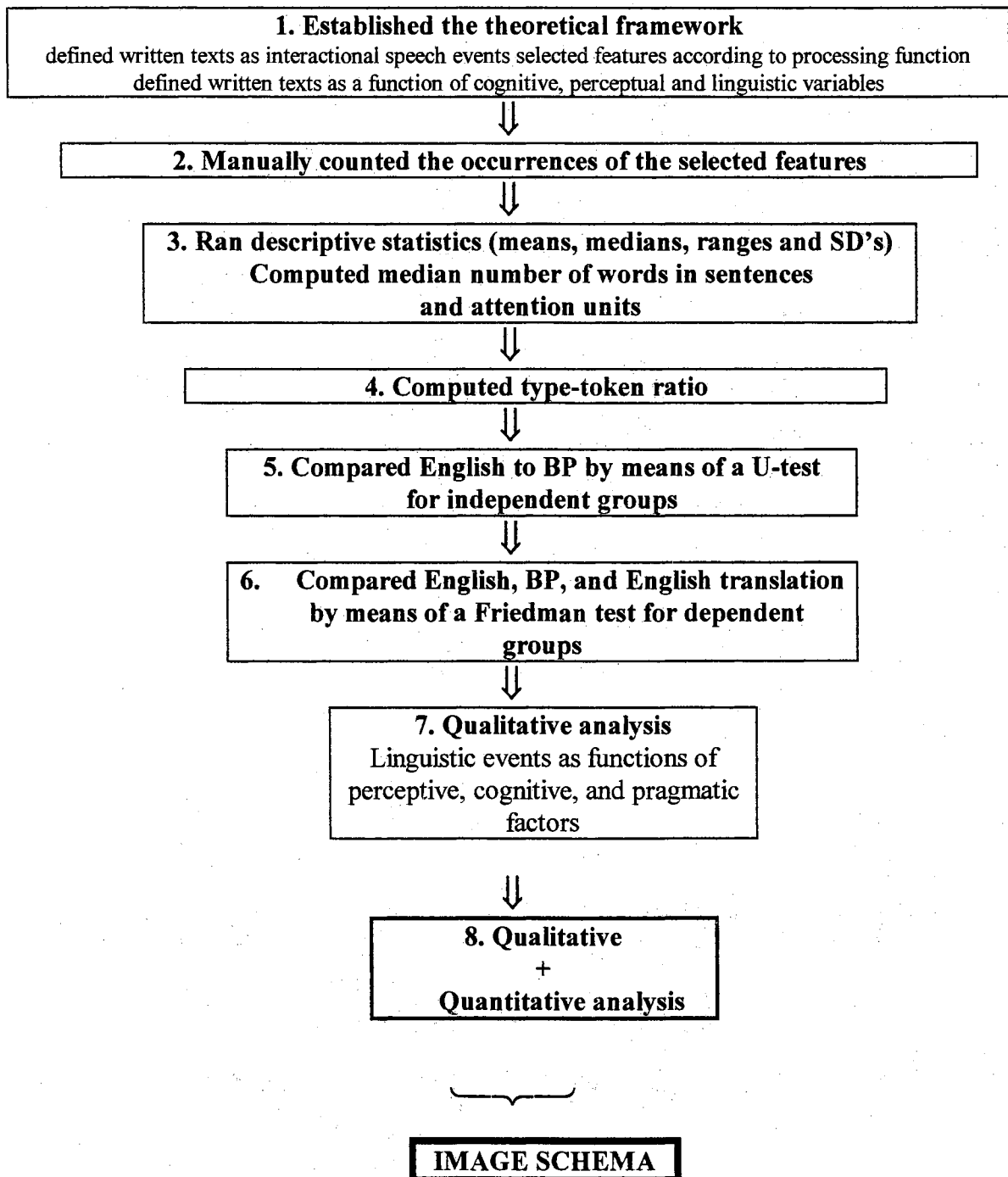


Figure 8: Getting at an image-schema of English and BP institutional discourse:
Methodological steps

comprehension: type-token ratio, number of content words and nominalizations, length of sentences and of attention units, number of independent clauses/sentences and adverbials of time and place. While attributes like a low type-token ratio may increase the ease with which readers construe an intended image schema, nominalizations may decrease.

Table 2 outlines the events the study analyzes. To facilitate the presentation and discussion of results, I clustered the attributes under three major variables: *organization of the lexicon* (type-token ratio and number of content words), *organization of syntax* (number of nominalizations, running words per sentence and per attention unit, and number of independent clauses/sentences), and *interaction with reality* (number of adverbials of time and place). Below I describe the specific measures.

TABLE 2
LINGUISTIC EVENTS

1.	ORGANIZATION OF THE LEXICON
	<ul style="list-style-type: none"> • Type-token ratio • Number of content words
2.	ORGANIZATION OF SYNTAX
	<ul style="list-style-type: none"> • Number of nominalizations • Sentence length • Length of attention unit • Number of independent clauses/sentences • Number of right branching constructions <ul style="list-style-type: none"> - two PPs in sequence - three PPs in sequence - four or more PPs in sequence - two or more PPs in sequence
3.	INTERACTION WITH REALITY
	<ul style="list-style-type: none"> • Number of locative and time adverbials

(1) *Organization of the lexicon.* This measure consists of a computation of the type-token ratio of words and the number of content words in the data. The type-token ratio reflects lexical variety and the number of content words reflects lexical density. Below I define and exemplify each measure.

The *type/token ratio* is defined as the number of different words (=the types) divided by the total number of words in that sample (=the tokens). For example, the paragraph below contains 27 tokens, but only 24 word types, given that the word **the** occurs three times and **of** twice (see in bold face):

- (1) Over the last decade, as understanding and aspirations with regard to protecting **the** environment evolved, society has become increasingly aware of **the** impacts of past human activity.

The type-token ratio is thus 24/27, or 0.89, indicating high lexical variety, and therefore, a low number of repeated words.

With regard to coding and tallying the types and tokens, every number, acronym, and abbreviation were considered tokens. In addition, words that belong to the same paradigm (e.g. production, produce, productivity; a, an) were counted as one type because paradigmatic relations are a form of repetition or cohesive ties (Chafe, 1994; Halliday, 1978; Johnstone, 1987). The excerpts below illustrate these coding criteria. Items that were not counted as types are underlined>, and items that are abbreviations, numbers and that were coded as types and tokens appear in bold face.

- (2) Shell E&P companies are already implementing a Safety Management System (SMS) as an extension of their existing Enhanced Safety Management Programme. (20 / 23= 0.87 ratio)

- (3) When the **Metropolitan Opera in New York City** opened its **1979-80** season in September, the occasion was carried nationwide on public television, made possible in part by a major grant from Texaco. (30/33=0.91 ratio)

Note also that compounds and proper names like New York City were counted as three tokens (in bold face). The symbol & was also tallied as a token. It takes time to conceptualize it.

Examples (1), (2), and (3) are instances that reveal a high lexical variety. The higher the ratio, the greater the lexical variety and the lower the number of repeated words. Though repetition may be disfavored in composition manuals and by folk attitude (Johnstone, 1987), according to existing research, lexical repetition is one of the structuring devices that brings to light the discourse focus of assertion (Johnstone; Tannen, 1988; Halliday and Hasan, 1976). Actually, Johnstone refers to repetition as a crucial cohesive device because it creates “shared language” and evokes “shared memory” (p.212). As such, it optimizes processing effort toward comprehension. Therefore, a chunk of discourse like the one in example (1) above could potentially cause comprehension constraints given its high lexical variety.

Furthermore, repetition of forms that enter in paradigmatic relations (e.g. to produce, production, producer, productivity) are facilitative because a first mention (e.g. to produce) activates the basic conceptual structure or cognitive domain. Thereafter, the rate of access to the PRODUC cognitive domain increases at each subsequent mention of a member that belongs to that same domain (e.g. production, producer, productivity). Likewise, the cost of activation of any member of that domain decreases (see Bower, Clark,

Lesgold, and Winzenz. 1969; Graf, Squire, & Mandler, 1984; Anderson, 1995 on the associative priming effect).

To investigate how the contrasted languages go about using lexical variety in terms of open and closed classes, I included a normalized frequency count of *content words* in the measure of lexical organization. This will allow us to see to which extent the corpora rely on the lexicon to cue meaning. In other words, it will provide us with a measure of lexical density (Halliday, 1989). According to Halliday, lexical density is revealed by the number of content words per running words in a text.

In the lexical density measure, every content word, even when repeated or member of a paradigm, were counted. Content words are those large classes of words to which new items can easily be added—nouns, verbs, adverbs and adjectives, as any introductory book to linguistics attests (e.g. O’Grady, Dobrovolsky, and Aronoff, 1993; Akmajian, Demers, Farmer, and Harnish, 1990). For example, in the excerpt from “Existing operations” (in Shell’s *The Test of Tommorrow*, p. 6), all words in bold face were coded and counted as content words:

“Determination of impact requires assessment of the fate of contaminants in the environment and understanding of their effects on the ecosystem. Combining emissions data with predicted or observed effects enables the operator to prioritise areas for emissions reduction.”

Normalizing the count, the number of content words for this example is 579:1000.

Similar to the count of types and tokens, numbers and possessives were counted as content words and compounds were treated as two content words (see in bold face).

Forms that were counted as content words are underlined:

Three months later, on December 8, 1979, Texaco's sponsorship of the Saturday matinee broadcasts live from the Met again its 40th consecutive year.

= (17 counts X 1000) : 23 running words = 739:1000

However, proper names like New York City or compounds like 1979-80 were counted as one single content word (see in bold face), given that they refer to one specific cognitive domain that is part of our concrete reality:

When the **Metropolitan Opera in New York City** opened its **1979-80** season in September, the occasion was carried nationwide on public television, made possible in part by a major grant from Texaco.

= (19 X 1000) : 33 = 576:1000

Content words were included in the study because (as opposed to functional words, which are linguistic proper,) they potentially cue meaning through semantic memory, and by relating to our psychological reality. That is, they enter in a net of semantic associations that are part of our life experiences and therefore tend to facilitate recall (Saliés, 1995a; Sumby, 1963). In this respect, content words we often use in everyday life (high frequency words) yield higher recall scores than those we rarely use (Sumby, 1963; Hall, 1954). The higher the number of associations they elicit, the lower the cost of activation (Noble, 1952). It is important to note, however, that when a net of rare content words enter in associations to construe the topic of a text, they may constrain comprehension.

Another good reason to include content words in this measure is that they generally leave subjects free to use top-down processing and function exclusively in the domain of meaning. That is, multiple synonyms, antonyms or collocations are cohesive ties as well

(Halliday and Hasan, 1976; Halliday, 1978; Johnstone, 1987). As such, like repetitions, they construe the topic.

To reiterate, a measure of lexical density tells us how texts in the corpora go about cueing meaning. That is, if they rely more on the lexicon or on grammar. If they rely more on the lexicon to cue meaning, their writers had to choose words carefully to avoid misunderstandings. For that matter, a measure of lexical density also reflects precision in word choice.

In brief, measures of lexical variety and lexical density indirectly point to the number of repetitions, synonyms, antonyms or semantic-related lexicon on which language users may rely to identify a referent or create meaning (Tyler, 1992; Halliday, 1989). By measuring lexical variety and lexical density, this study assessed how the different research conditions organized and used their lexicon to optimize comprehension.

(2) *Organization of syntax*. This measure consists of a manual count of the number of running words per sentence and per attention unit (stretches of language between punctuation marks), nominalizations, independent clauses/sentences, and right-branching constructions such as prepositional phrases in sequences. Its relevance derives from the figure-ground asymmetry. A more elaborate syntax signals to readers what is ground (dependent constructions) and what is figure (the foregrounded information or autonomous entity), structuring discourse. Next, I examine each level of the syntactic construction category in turn.

Sentence length. This measure consists of counting the number of running words per sentence and averaging it for the total number of sentences in each text to get at a mean length of sentences in each language. A sentence, the major unit of writing (Chafe

and Danielewicz, 1987), is understood as the words between two orthographic periods, question marks or exclamation marks (Dantas-Whitney and Grabe, 1991).

In counting the number of running words in a sentence, clitics, acronyms, abbreviations, possessives, and numbers were treated as one word or token. Compound words, however, were treated as two tokens. These cases appear in bold face and are underlined in the excerpts below:

- (1) “At its **R&D** Center, known as Cenpes, Petrobras maintains an automotive emission control laboratory to subsidize work on improved fuel and engine quality.” (Petrobras, Environment, Quality and Safety, “Technology at the Service of Better-quality of Living”)
= 25 words; R&D was counted as three tokens—Research and Development
- “O Centro de Pesquisas da Petrobras, através dos laboratórios de emissões veiculares, auxilia a gerência da qualidade dos combustíveis e dos motores.” (Petrobras, *Meio Ambiente, Qualidade e Segurança*, “Petroleo, um velho conhecido”)
= 22 words
- (2) “Petrobras Distribuidora **S.A.**, which goes by the trade name **BR**, is the group’s subsidiary for **oil-product** distribution. **BR** markets some **350** items, mainly fuels and lubricants consumed in all corners of Brazil.” (Petrobras, Environment, Quality and Safety, “Secure Delivery of Top-rate Oil Products”)
= 34 words; S.A. was counted as two tokens; BR was counted as one
- “A Petrobras Distribuidora **S.A. (BR)** é a subsidiária que desempenha esta atividade, oferece ao mercado cerca de **350** itens, onde predominam combustíveis e lubrificantes que devem chegar a todas a regiões do Brasil. (Petrobras, *Meio Ambiente, Qualidade e Segurança*, “Distribuição de produtos com qualidade e segurança”)
= 34 words; S.A. was counted as two tokens—*sociedade anônima*
- (3) “O petróleo é um velho conhecido da humanidade, que começou a utilizá-**lo** há muitos séculos...” (Petrobras, *Environment, Quality and Safety*, “Oil, a friend from way back”)
= 16 words
- “Oil has for centuries been **humankind’s** friend, ever since people first learned to make use of small amounts of this material...” (Petrobras, *Meio Ambiente, Qualidade e Segurança*, “Petroleo, um velho conhecido”)
= 21 words

- (4) “Shell UK’s Better Britain campaign has supported **community-based** environmental projects for the past 30 years.” (*Shell and the Environment*, “Case study—Environmental conservation” p.852).
 =16 words; UK’s was treated as one token as well as EUA and other abbreviations for geographical references.

Given that existing studies show that writers possess an intuitive concept of “normal sentence length,” sentence length appears to be an important attribute of the text schema. According to Chafe and Danielewicz (1987), it generally averages 24 words per sentence in written English. Unfortunately, the type of discourse for which this normal length has been computed was not indicated. Furthermore, as I mentioned before, sentence length seems to be language specific. This study investigates that fact.

Length of attention unit. Following Chafe (1980;1985;1988; 1994) and Chafe and Danielewicz (1987), I assume in this study that “written language has a covert prosody which is analogous to that of spoken language: both writers and readers assign pitch, stress, and pauses to language as they write and read it” (Chafe and Danielewicz, p.110). This covert prosody establishes boundaries similar to what has been traditionally known as intonation units in spoken discourse and it approximately corresponds to “stretches of language between punctuation marks” (p.110). Chafe (1988, 1994) refers to those units in written language as “punctuation units. In this dissertation, because such boundaries seem to reflect the extent to which language users can focus attention or keep an amount of information in active consciousness, I call them “*attention units*.”

Therefore, the category *length of attention unit* consists of counting the number of running words between punctuation marks and averaging it out over the total number of attention units in each text. Below are five examples of what I counted as “attention

units.” They were extracted from *Exxon and the Environment* (1990, p.106) and appear in sequence, forming one continuous stretch of discourse:

- | | |
|--|-----------|
| (1) ...you will love watching the terns return to Exxon Company, | =10 words |
| (2) USA's Grand Isle Gas Plant. | = 5 words |
| (3) Each April, | = 3 words |
| (4) right on schedule, | =3 words |
| (5) thousands of small gull-like birds flock to this small island in the Gulf of Mexico. | |
| =16 words | |

The average number of words per attention unit in the passage above is

$$\frac{(10+5+3+3+16)}{\text{total number of running words}} : \frac{5}{\text{number of attention units}} = 7.4 \text{ Length}$$

Again, abbreviations, acronyms, possessives, clitics, and numbers were counted as one token. Compound words were treated as two tokens. Apostrophes and hyphens were disregarded as marks of punctuation for this analysis.

According to Chafe and Danielewicz (1987), the mean length of attention units adds up to 9 words in written academic English discourse. Chafe (1988) found attention units that varied from 10 to 13 words in published academic discourse. Both found that most often these units comprise a single clause, but that they may carry only a phrase of some kind or even fragments (Chafe and Danielewicz, 1987) as in examples number three, four, and five. These researchers theorize that the limited size of attention units (only moderately higher than those found in conversations—6 words per unit) may reflect writers' concern about their audience's limited capacity of focal consciousness. If writers expand the units too much, they may impact ease of comprehension, and ultimately communication, their main objective.

In addition, though some readers may not assign stress and pauses as they read, and despite differences writers may have in punctuation style, the regularity with which attention units of 9-11 words appear in academic written English potentially indicates their ecological validity as a measure of production and comprehension. Chafe (1988) elaborates this idea by demonstrating that most writing in his study used spoken-like punctuation units, respecting the prosody of written language. According to him, length of attention unit tells us to which “degree the punctuation of a particular piece of writing accords with the prosody of spoken language” (p. 408). If this is true, most attention units, similarly to intonation units, respect “the one new idea constraint” (Chafe, 1994), and reveal the flow of ideas through writers’ and readers’ minds.

Nominalizations. The forms coded in this tally are complex high order schemas that construe the process profiled by the verb as regions in abstract domains (Langacker, 1991). In other words, they are events or states that have been reified to acquire conceptual persistence and appear in other events or states (Chafe, 1994). For example, importation (= *importação* in Brazilian Portuguese) profiles the process to import (= *importar*) as a region, a thing in the conceptual domain import when added of the suffix -ion (= *ção*). The same is true of arrival (= *chegada*). Other examples from the data include production (= *produção*), exploration (= *exploração*), planning (= *planejamento*), and prospecting (= *prospecção*), housing (= *habitação*), and incidence (= *incidência*).

Both in English and BP nominalizations derive from suffixation. The English derivational morphemes used in this count were: *-ation, -al, -ance, -ion, -ment, -y, and -ty* (meaning the act of; e.g. *execution, refusal, performance, conversion, measurement*); *-acy and -ence*, and (meaning the result of an action; e.g. *accuracy,*

independence); -er, -or (a thing or person who performs an action; e.g. *reporter*, *operator*, *driller*).

Equivalent derivational morphemes were coded in BP: *-ança*, *-ância*, *-ença*, *-ência* (meaning the result of an action; e.g. *permanência* [=permanence], *cobrança* [=collection of debts]; *concorrência* [=competition]); *-ante*, *-ente*, *-inte*, or (a thing or person who performs an action; e.g. *operador* [=a male operator]; *produtor* [=producer]; *estudante* [=student]; *ouvinte* [=hearer]), *-ção* and *-são* (result of an action; e.g. *perfuração* [drilling operations]; *contaminação* [=pollution]; *extensão* [=extension]), and *-mento* (action or result of an action or instrument of action; e.g. *equipamento* [=equipment]; *tratamento* [=treatment]).

Only cases that could be found both in English and Portuguese, and that were produced from verbs were coded and tallied. For example, words like *activity*, *difficulty*, *capacity*, *acceptability* derive from adjectives in English. Therefore, they were not coded. The same holds true for BP nominalizations. Words like *atividade*, *dificuldade*, *capacidade* and *aceitabilidade* were not produced from verbs. Therefore, they were not tallied.

In addition, the count also included (1) standard items or well-established technical concepts such as *exploration* (=produção), *drilling operations* (=perfuração), and *production* (=produção); (2) plural inflection of BP nominalizations (e.g. *desenvolvimentos*; *perfurações*; *concorrências*); (3) *-ing* endings that necessarily become a nominalization in BP (e.g. *Seven decades of prospecting* [*prospecção*] *came true*); (4) words with these affixes regardless of whether the process was discourse old.

For example, in excerpt number 1 below, there are three nominalizations (see in bold face). If we normalize that count using the total number of words in the paragraph, there are (3 X 1000) : 59 running words. That gives us 51:1000 nominalizations.

- (1) The costs of preventing, controlling or reducing **emissions**, discharges and wastes at operating facilities are regarded as a normal part of conducting business. These amounts are thus not accounted for separately. However, it is estimated—based on **allocations** and managerial **judgment**—that Group companies in Europe and North America incurred amounts of the order of £950 million in 1994.

Even nominalizations that appeared in the place of a modifier, in head-modifier relations, were coded and tallied because they form an NP. In BP the modifier becomes the head-nominalization:

- (2) *Most important, there is no shortage of coal, the basic ingredient for the **gasification** process. Coal is our nation's most abundant energy resource.*

O mais importante é que não há escassez de carvão, o ingrediente básico do processo de **gasificação**. O carvão é o recurso energético mais abundante do país.

Next, I show in bold face other examples of expressions in the corpora that were coded and tallied as nominalizations:

- (3) *They range from jungle depths to desert dunes, from Ameridian reserves to **development** complexes.*

Vão do coração da floresta até as dunas desertas; das reservas indígenas até os polos de **desenvolvimento**.

- (4) *Shell has a 50 per cent share in Wavin, Europe's largest **processor** and **recycler** of plastic products.*

A Shell possui 50% das ações da Wavin, a maior empresa de **reciclagem e processamento** de plástico da Europa.

- (5) *Reducing oil **pollution** from offshore **production** facilities or tankers, whether from routine **operations** or from accidents, demands **adherence** to high **management** standards within the industry. **Prevention** is accepted as the best of all oil spill responses, and Shell companies put considerable effort into training and establishing high standards of **operations** aimed at reducing the risk of incidents occurring.*

Reduzir a **poluição** causada pela **produção** de óleo em plataformas marítimas ou por petroleiros, seja por consequência de **operações** rotineiras ou de acidentes, exige alto padrão de controle na indústria de petróleo. A **prevenção** é aceita como a melhor **solução** para **vazamentos** de óleo e as empresas Shell dedicam-se com esforço a treinar e estabelecer um alto padrão de **operação**, visando a **redução** de incidentes.

- (6) Shell forestry **operations** are limited mainly to tree **plantations** for supplying raw materials to wood processing industries.

O **reflorestamento** realizado pela Shell se limita principalmente à **plantação** de árvores que servem como matéria-prima para a indústria madeireira.

- (7) *For instance, environmental **assessments** and **consultations** with interest parties, including local people, are essential parts of the **planning** process.*

Por exemplo, **avaliações** e consultas com as partes interessadas, inclusive a **população** local, constituem partes essenciais do processo de **planejamento**.

- (8) *This trial programme rapidly demonstrated the value of sending “**Inspectors**” aboard the ships.*

Este processo experimental rapidamente demonstrou o valor de mandar “**inspetores**” a bordo dos navios.

Note that in example number 5, there are three nominalizations with the *-ing* suffix.

Those forms, however, have not been included in the count because in BP they do not necessarily become nominalizations. In this case they became to-infinitives. While in English the passage has (6 X 1000): 58 nominalizations (= 103:1000), in BP the same passage yields (8 X 1000): 66 nominalizations (= 121:1000).

Finally, nominalizations that are part of proper names and refer to physical entities were not tallied (e.g. *National Department Production*) while those that referred to

action plans were coded (e.g. *Safety Management Program; Environmental Management System*)

Why code and tally nominalizations? Nominalizations pack and objectify information that could otherwise be spelled out in a typical action-chain. In that they hide semantic relations of a typical action-chain, they may be more costly to process. Take example number 13:

(13) **The improvement of gas emission by Petrobras pleased public opinion.**

Though the lexical nominal (=improvement) is a trajector (TR), it fits marginally in the action-chain metaphor as the source of energy. Actually, it is much easier to construe *Petrobras* as the source of energy. The difficulty derives from the fact that while *Petrobras* is a prototypical NP, *improvement* is a less typical member of the NP category (see Givón, 1993). Indeed, it reifies the process profiled by the verb *to improve* that has as its source of energy *Petrobras*. Furthermore, it hides the cause-effect relation between the argument and the predicate. If we unpack the nominalization *improvement*, this cause-effect relation stands out, *Petrobras* becomes the source of energy, and the verbal process is profiled:

Petrobras please public opinion because it improved gas emission

When cases similar to the example happen with high frequency, many semantic relations become hidden, and comprehension may take more time or even fail.

In addition, if readers of the brochures selected for the data were not people from the oil industry, many of them would probably not know that *gas emission* means *to*

control substances that pollute the air. Thus, if we unpack the second nominalization—*emission*, sentence (13) can still be clearer than its second version. The final result would be

Petrobras pleased public opinion because it improved the control of substances that pollute the air.

The cues to retrieve the underlined speech act are not in the nominalization itself, but are part of readers' background knowledge and familiarity with the oil industry jargon. In the case of the intended audience, this type of nominalization brings cognitive gains. Most readers know the meaning of gas emission. The full mention of the process would cause an unjustified processing effort. The reification is, then, cognitively positive.

Likewise, there are nominalizations that structure information in discourse (Watson and Olson, 1987; Halliday, 1987; 1989), and, consequently, bring positive cognitive effects. When nominalizations structure information, they allow further reference and development of the topic. For instance, I can refer back to the abstract domain *improvement* in example number 13 through the anaphora *it*.

To resolve the contrasting roles nominalizations take in discourse—source of ambiguity, distancing device, and organizer of information—we need to call on Relevance Theory (Sperber and Wilson, 1986). If the information the nominalizations convey is part of the context or given (semi-active) information, their presence represents the smallest processing cost. On the other hand, if that information is decontextualized, the cost of processing increases. Nominal forms become, then, a source of ambiguity and misunderstandings, and violate The Principle of Relevance as outlined by Sperber and

Wilson. I refer the reader to Chapter II for a detailed review of studies that dealt with nominalizations and with Relevance Theory.

Right-branching constructions. This category consists of a manual count of the number of prepositional phrases that appear in sequence (two or more juxtaposed PPs). I tallied sequences of two juxtaposed PPs, sequences of three juxtaposed PPs, sequences of four or more juxtaposed PPs, as well as provided a total number of instances in which two or more juxtaposed PPs appeared in sequence. A prepositional phrase (PP) is syntactically defined as a preposition (P) plus a noun-phrase (NP) as in the expressions in bold face:

(1) **Over the last decade**, as understanding and aspirations with regard to protecting the environment evolved, society has become increasingly aware **of the impacts of past human activity**. (Shell, 1993, p.8)

(2) *Na última década, à medida que o entendimento e o desejo de proteger o meio ambiente aumentou, a sociedade tornou-se cada vez mais consciente dos impactos causados **pelas atividades do homem no passado**.* (my translation)

In example number one above, there is only one case of two juxtaposed PPs:

of the impacts
of past human activities

In example number two, the BP variant of the passage in example number 1, there is one sequence of three juxtaposed PPs:

pelas atividades
do homem
no passado

Examples number three and four illustrate a sequence of four or more juxtaposed PPs:

- (3) Also, in 1979, Texaco began site preparation
**for field tests
of a new recovery process
at the Company's oil shale property
in Utah.**

- (4) A província (*the site*)
de gás (*of gas*)
de Juruá, (*of Juruá*)
no município (*in the town*)
de Caruaru, (*of Caruaru*)
acerca de 750 quilômetros (*about 750 kilometers*)
de Manaus. (*from Manaus*)

The distribution of these constructions is shown in terms of occurrences per thousand words, similar to other constructions the study selected to analyze. In other words, the count of juxtaposed prepositional phrases is multiplied by 1000 and divided per the total number of running words. Other cases that have been coded and tallied for this measure are illustrated below.

Sequences of two juxtaposed PPs:

- (1) Operations provide compelling evidence **of the environmental vitality of the Sound and the Gulf.**
- (2) All work recommended **by the Joint Technical Advisory Group** and by **the US Coastal Guard** was completed.
- (3) The company is also a market leader **in the manufacture of returnable plastic crates.**

Sequences of three juxtaposed PPs:

- (4) Fahrenheit and pressures **of thousands of pounds per square inch.**
- (5) Determination of impact requires assessment **of the fate of contaminants in the environment** and understanding of their effects on the ecosystem.
- (6) That Group companies in Europe and North America incurred amounts **of the order of £ 950 million in 1994.**

Note that PPs that appeared coordinated by the conjoiner and (examples number 2 and 4) were counted as in sequence. Structurally they are in sequence. Also, note that example

number 5 has a sequence of two juxtaposed PPs right after the first sequence of four (*understanding of their effects on the ecosystem*).

Sequences of four juxtaposed PPs:

- (7) As a contribution **to the debate on the environmental and social acceptability of tree plantations for wood production**, SIPC together with the World Wide Fund for Nature (WWF), has produced the Tree Plantation Review.
- (8) It represents the longest continuous sponsorship **of a radio program by a single company in the history of broadcasting**.

If we consider the eight examples above as a being parts of a single text, they constitute nine instances of two or more juxtaposed PPs (the total number of juxtaposed PPs). That is, a sum of all instances in which I encountered two, three, four or more juxtaposed PPs. To compute a normalized count of that number, we multiply the number of two or more juxtaposed PPs per 1000 and divide by the total number of words:

$$(9 \times 1000) : 154 = 58:1000$$

Last, prepositional phrases that function as adverbials of time and place (e.g. *over the last decade* or *in the past*) were coded both under PPs and under the category adverbial phrases (see next sub-heading) because structurally they are PPs and functionally AdvPs.

On the other hand, PPs that appeared as complements of verbs (indirect objects) were excluded from the tally:

- (1) The current policy calls **for continuous improvement**.
- (2) Shell companies have access **to the extensive clean-up equipment** provided by the global and regional industry-managed Tier.
- (3) The resources mobilized depends **on the severity** of the spill.
- (4) Shell forestry operations are limited mainly **to tree plantations** for supplying raw materials to wood processing industries.

The PPs in these cases are part of the verb sub-categorization and together with the verb constitute a single mental structure. Verb sub-categorization is one of the reasons I had to count only juxtaposed PPs. Structurally, Brazilian Portuguese is a head-initial language that favors the appearance of PPs or post-modifiers. Note the example below. BP shows three juxtaposed PPs whereas English shows two. This difference is a direct consequence of word order:

society has become increasingly aware **of the impacts of past human activity**.

*a sociedade tornou-se cada vez mais consciente **dos impactos** causados **pelas atividades do homem no passado***

While past human activity appears as a head-modifier relation in English, in BP it appears as a head-complement relation—*atividades do homem no passado*.

An additional reason for counting only juxtaposed PPs rests on the cognitive nature of PPs. According to CG, a PP is an atemporal relation between a conceptually dependent structure and an autonomous structure. However, the e-site of the autonomous structure appears within the PP. In other words, a salient substructure of the complement elaborates the head. For example, in the head-complement relation *application of technology*, the word *technology* elaborates the domain *application*. Therefore, PPs are profile determinants.

Subsequent elaborations of a head relates positively to the amount of time it takes to construe the final composite structure. Readers may even fail to construe the image intended by the writer if these sequences are too lengthy. Sequences of PPs as negative

cognitive effects may be especially true of English, a language with few morphological trappings to tie components of the TR-LM relation.

Furthermore, from a functional perspective, prepositional phrases normally increase the size or the number of attention units (Chafe and Danielewicz, 1987) and, ultimately, they increase the length of sentences. If PPs increase the number of words readers need to keep under the focus of attention (Cuetos and Mitchell, 1988, and Hoover, 1992; for further detail see Chapters III and IV), they may overload working memory and impact comprehension. Thus, when juxtaposed, PPs are likely to bring cognitive strain. The longer the sequence the greater the cognitive effort to process it.

In brief, the right-branching constructions coded and tallied in the data are juxtaposed PPs—conceptually dependent structures that focus the foregrounded entity. Within the category, I counted two, three, and four or more juxtaposed PPs. In addition, I also tallied the number of two or more juxtaposed PP that appeared in the data. Structures with the same characteristics that I described above were coded for the Brazilian Portuguese data.

Independent constructions. Independent constructions foreground information and are typified by clauses/sentences like Our emphasis on risk reduction and emergency response capability reflects an organization wide dedication (*Exxon and the Environment*, 1993, p.28) or The current policy calls for continuous improvement (Shell, 1993, p.3). These clauses designate the same process profiled by their verbs, being autonomous conceptually. As Langacker (1991) says, they require nothing further. Similar counterparts for this count were compiled for the Brazilian Portuguese data.

(3) *Interaction with reality*. This measure consists of a manual count of the number of adverbial phrases of time and place. Like in other instances, I computed the number of occurrences per thousand words.

Specifically, I coded and tallied time and location expressions that function as adjuncts. Some are single words, others are noun phrases, prepositional phrases or even adverbial clauses (see Greenbaum and Quirk, 1990). In spoken language, these forms would generally appear in a separate intonation unit. For example, in the sentence

- (1) **Three months later, on December, 1979**, Texaco's sponsorship of the Saturday matinee broadcasts live from the Met began its 40th consecutive year" (*Texaco 1979 Annual Report*, p. 29)

the expressions in bold face were counted as two time adverbial phrases (three months later and on December 1979). These adjuncts somehow relate to the topic or subject of the sentence and indicate involvement with the concrete reality or with the subject matter (see Chafe and Danielewicz, 1987).

Other examples of adverbials of time and place tallied in the corpus include:

- (2) The environment in **Prince William Sound and the Gulf of Alaska** is undergoing a rapid recovery from the 1989 oil spill (*Exxon Corp. 1990 Annual Report*, p.28)
- (3) Resources committed to the effort **in 1990** included approximately 1,000 people..." (p.28)

Cases in which places and different points in time were conjoined were considered two instances. Example number two above illustrates that criterion. In it, there are two instances (in Prince William Sound and the Gulf of Alaska) of locative adverbials.

The inclusion of this measure in the study is based on the importance of concrete reality or world knowledge to the process of meaning formation. Frequently, because discourse fails to bring explicit mentions about time and place of events, readers access that information through context or general knowledge (which is reader conscious, but semi-active). If the context and readers' world knowledge are appropriate, there is no processing constraint. The information is retrievable. However, to activate information that is displaced or that is part of readers' scripts, there is a cost. Also, if the implicature is weak or world knowledge fails, language that explicitly refers to concrete reality becomes a processing facilitator.

In addition, if a given space builder (=locative and time adverbials) is explicitly mentioned in discourse, the fact that it is said makes it pragmatically salient or prominent and improves the accessibility of entities that are part of that space.

To wrap up, this study investigates how BP, English, and English translation institutional expository texts organize and use the lexicon, syntax, and how the three research conditions interact with concrete reality by means of locative and time adverbials. The measures the study uses to do so are the type-token ratio, the normalized count of content words, mean length of sentence and attention units, a normalized count of nominalizations, two or more juxtaposed PPs, independent constructions, and time and locative adverbials. In addition, it qualitatively analyzes how these devices appear in the corpus of each research conditions in the light of cognitive grammar, pragmatics, cognitive, and perceptual factors. As the qualitative analyses proceeds, the study also entertains the presence of word-order effects by observing head-modifier relations, inflections, SV agreement, and right branching constructions such as non-finite clauses.

The co-occurrence of these linguistic attributes may relate to effort of processing as well as to an image-schema of what I called the COMMUNICATIVE TEXT. As I mentioned in chapter one and outlined in the introduction of this chapter, my research goals are

1. To analyze the distribution of the 8 linguistic devices across the 20 institutional expository texts in BP, English translation, and English and propose a schema for the COMMUNICATIVE TEXT category in each research condition.
2. To explain how the linguistic attributes in each research condition relate to effort of processing.
3. To provide a cognitive rationale for discourse translation.
4. To psychologically explain differences between English and BP, contributing with a theory of discourse analysis that combine competence with performance.

In the next chapter, I describe and discuss my findings, presenting quantitative and qualitative results for each research condition, comparing English to BP, comparing the three research conditions, and explaining discourse translation cognitively.

CHAPTER VII

TEXTS AS IMAGE-SCHEMAS: THE COMMUNICATIVE TEXT CATEGORY

Overview

In this chapter, I present the results of the study described in Chapter VI and designed to identify typical attributes of English, Brazilian Portuguese, and English translation institutional discourses. To do that, I present frequency counts and normalized counts for each category across the research conditions. Qualitative analyses of the attributes follow immediately thereafter. In addition, I compare the results of the three research conditions and provide a cognitive rationale for discourse translation.

Since the texts used in the study failed to yield a normal distribution with regards to the attributes analyzed, the comparisons among the groups will take the medians as the measure of central tendency, and ranges as the measure of variability.

English

The English “communicative text category” activates knowledge of a series of events associated with English texts. To find out which events are typical of English or have high representativeness, I examined the frequency of occurrence of the selected attributes. The present section will consider each of these events in turn.

Sentence Length

Table 3 and 4 illustrate sentence length in English institutional written discourse. Table 3 displays the median frequency of words per sentence and the variability of the data

from the central tendency (Median=20; range=53). The most typically used sentences have between 14 and 24 words. However, if it were not for the extreme score of 55, the median would be lower than 20. Many more sentences group at the lower end of the range (minimum of two and maximum of 55 words), reflecting a positively skewed distribution. A further look at the frequency of use of sentences with different lengths confirms that tendency. Out of 301 sentences, 151 (50.1%) have between 14 and 24 words, 57 (18.9%) have between thirteen and two words, 73 (24.4%) varied between 25 and 35 words, and only 20 (6.6%) have more than 36 words. To determine the intervals in Table 3, I looked at the frequency counts. Sentences that occurred moderately ($4 \leq f \leq 9$) appear first; sentences that occurred more than 10 times or with high frequency ($10 \leq f \leq 19$) appear second; sentences that also occurred moderately ($4 \leq f \leq 9$) appear third; and sentences that rarely occurred ($3 \leq f \leq 1$) appear last.

TABLE 3
SENTENCE LENGTH IN ENGLISH
n=301

Descriptive Statistics		Number of words	<i>f</i>	%
Median	20			
Max. value	55	02-13	57	18.9%
Min. value	02	14-24	151	50.1%
Range	53	25-35	73	24.4%
		36-55	20	6.6%

Table 4 rank orders the most used sentence lengths in the sample as well as shows the percentages and frequency of use as a function of the number of words. The most widely used sentence length is 15 words per sentence, immediately followed by sentences

with 20 and 21 words. These results reveal a tendency English institutional discourse has to keep sentences within the bounds of 24 words.

TABLE 4
 MOST USED SENTENCE LENGTH IN ENGLISH
 number of sentences = 301
 RANK ORDERED

Rank	Length	%	<i>f</i>
1	15	6.3	19
2	21	6.0	18
2	20	6.0	18
3	23	5.0	15
4	22	4.3	13
4	19	4.3	13
5	17	4.0	12
5	16	4.0	12
5	26	4.0	12
6	18	3.7	11
7	24	3.3	10
7	24	3.3	10

Figure 9 depicts the tendency. The bars represent intervals as listed in Table 3. That is, the first bar stands for sentences that have from two to thirteen words and that appeared in the English data 57 times. The second bar, my main focus of interest, represents sentences that have from fourteen to 24 words. These appeared in the English data 151 times constituting 50.1% of the overall sentences.

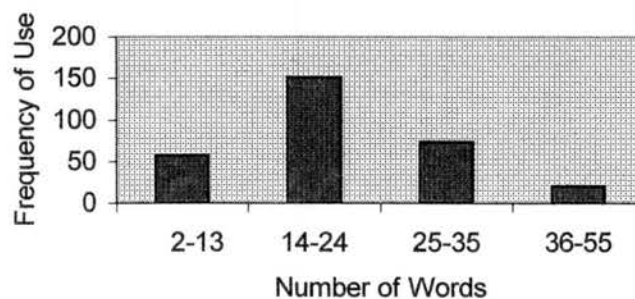


Figure 9. Sentence length in English.

Most sentences in English institutional discourse have up to 24 words.

Length of attention units

The distribution of words per attention unit is represented in Table 5. Compared to sentence length, attention units reveal a greater tendency to cluster around the median number of words per unit (Median=9; Range=41). The most typically used sizes of attention unit in the English data has from 3 to 9 words. However, the data is negatively skewed because of extremes. Attention units with two words appear 52 times in the data, representing 8.5% of the total 611 units. If we eliminate units that were used with low frequency ($f < 22$), the median will be higher.

To determine the intervals in Table 5, I separated the extremes at both ends of the scale (attention units with less than two words and attention units with more than 25 words) and examined the frequency counts. Attention units that occurred with high frequency ($30 \leq f \leq 45$) have from three to nine words; attention units that occurred with moderate-high ($17 \leq f \leq 29$) frequency have from ten to seventeen words; and attention

units that occurred with low frequency ($6 \leq f \leq 10$) have from eighteen to twenty-four words.

TABLE 5
LENGTH OF ATTENTION UNITS IN ENGLISH
n=611

Descriptive Statistics		words per unit	<i>f</i>	%
Median	9	1	21	3.4
Max. value	42	2	52	8.5
Min. value	1	3-9	248	41
Range	41	10-17	188	31
		18-24	72	12
		25-42	29	4.7

Table 6 lends further support to those results rank-ordering the most used attention units according to number of words and showing the units relative frequency as percents of the overall corpus. While attention units with 6 words occurred 39 times (6.4%), the

TABLE 6
MOST USED ATTENTION UNITS IN ENGLISH ACCORDING TO LENGTH
number of units = 611
RANK ORDERED

Rank	Length	<i>f</i>	%
1	2	52	8.5
2	4	45	7.4
3	6	39	6.4
4	3	37	6.1
5	7	33	5.4
5	9	33	5.4
6	8	31	5.1
7	10	30	4.9
7	5	30	4.9
8	15	29	4.7
9	11	26	4.3
10	16	22	3.6

median (=9) occurred 33 times (5.4%). The relevance of these results rests on the consistency with which the units are kept within the bounds of nine words (52%).

Figure 10 illustrates the percentage distribution of attention units as a function of the number of words. The bars follow the intervals indicated in Table 5. For example, the third bar represents attention units with three to nine words and they consist 41% of the total number of units; the last bar represents units with 25 to 42 words and they consist 4.7% of the total number of units in English institutional discourse. The representativeness of the nine-word constraint for length of attention units stands out.

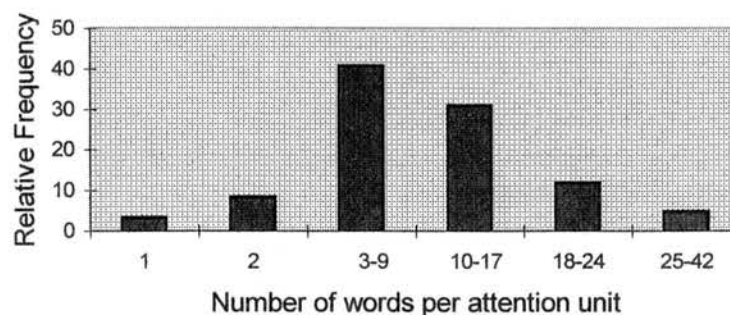


Figure 10. Relative frequency of attention units as a function of number of words in English.

Type-token ratio

Table 7 portrays the ratio of unique words (types) to the total number of words (tokens) in the English data across texts as well as the absolute and relative frequency with which the ratio occurred in the corpus. Descriptive statistics illustrate the somewhat even distribution of the ratio across texts (Median=0.55; Range=0.34). Though ratios range from 0.40 to 0.74, typically they stay above the fifties. This is true of 65% of texts.

TABLE 7
 DISTRIBUTION OF THE TYPE-TOKEN RATIO ACROSS ENGLISH TEXTS
 n=20 texts

Descriptive Statistics		Type-token	# Texts	%
Median	0.55	0.40-0.47	7	35
Max. Value	0.74	0.52-0.58	7	35
Min. Value	0.40	0.60-0.66	4	20
Range	0.34	0.70-0.74	2	10
Overall ratio for the corpus as a whole		.50		

The intervals in Table 7 reflect the fact that most texts in the sample have type-token ratios above the fifties. First, I report how many texts have ratios in the forties; second how many texts have ratios in the fifties; third, those that stay in the sixties; and finally those that are in the seventies.

Furthermore, at least half of the words in the corpus are different types, as indicated by the overall type-token ratio for the whole corpus (3168 : 6398 or .50). Institutional discourse in English displays a wide range of lexical variety.

Content words

The type-token ratio, however, does not inform us about how precisely writers choose words in the English corpus. To assess lexical density, I counted the number of content words expressed per 1,000 words in the sample. Table 8 displays the results. In each thousand words in the English data, there are 611 content words. That number is also the median of the distribution. Actually, out of twenty texts (n=14), only six fail to have more than 60% content words.

TABLE 8
 NUMBER OF CONTENT WORDS PER 1,000 WORDS IN ENGLISH TEXTS
 n=20 texts

Descriptive Statistics	content words
Median	611
Max. Value	682
Min. Value	554
Range	128
<hr/>	
TOTAL NUMBER OF OCCURRENCES:1000 FOR THE CORPUS	611

Figure 11 illustrates how nouns, verbs, adjectives and adverbs have the highest percentage in the total number of running words, representing (62%) of the total words in the corpus or 611 words out of a thousand words.

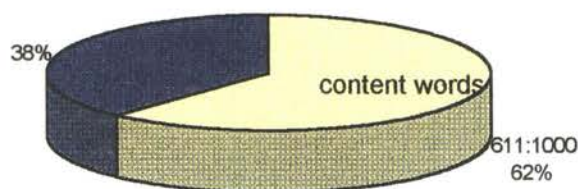


Figure 11. Ratio of content words in institutional English discourse.

Therefore, English institutional discourse also displays high lexical density. That is, the number of lexical items per running words reflects a greater reliance on the lexicon than on grammar to convey meaning. Or still, it reveals that most words in the English

corpus enter in open contrasts to pack information. If meaning is most expressed by means of the lexicon, words need to be precisely chosen to avoid misunderstandings. In this sense, lexical density in English reflects a precise choice of words for the part of the writer.

Nominalizations

In addition to choosing words precisely, English writers in the corpus draw on nominalizations to allow for further reference and development of discourse. The overall frequency of occurrence of nominalizations in the whole corpus was 61:1000. Table 9 outlines the uneven distribution of nominalizations per 1,000 words across the twenty English texts in the study (minimum of 12:1000 and maximum of 133:1000). There is no single number that stands out as the most typical, but a variety of numbers that evenly spread out from the median (Median=58; range 121).

TABLE 9
NUMBER OF NOMINALIZATIONS PER 1,000 WORDS IN ENGLISH TEXTS
n=20 texts

Descriptive Statistics		MOST FREQUENT NUMBER OF NOMINALIZATIONS	# Texts	%
Median	58	12-35	3	15
Max. value	133	40-45	5	25
Min. value	12	50-67	4	20
Range	121	76-86	4	20
		94-97	2	10
		106-133	2	10
TOTAL OCCURRENCES IN THE CORPUS:1000		61		

* NOTE: Totals may not add up due to rounding.

The only number that appears twice is 41:1000 and a high number of cases (45%; n=9) concentrate at the center of the distribution, displaying from 40 to 67 nominalizations per thousand words.

The intervals in Table 9 as well as their relative frequencies of occurrence are illustrated in Figure 12. The center of the distribution stands out (n=9; 45%). Furthermore, a great number of texts in the English corpus have from 40 to 86 nominalizations in each 1,000 words.

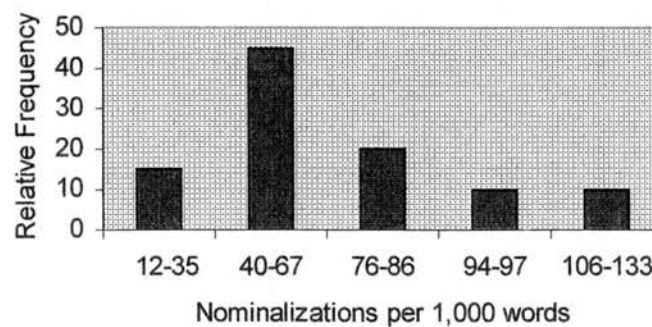


Figure 12. Relative Frequency of Nominalizations per 1,000 words in English discourse.

Sequences of Prepositional phrases

Though English discourse uses nominalizations somewhat frequently to develop discourse, it uses prepositional phrases with low frequency. Table 10 shows the frequency of two or more juxtaposed prepositional phrases in the English data per thousand words. The table also outlines observations for two, three, and four or more juxtaposed PPs. Results reveal that in each thousand words we find 17 sequences of two or more PPs.

Out of these 17 sequences, 13 consist of two PPs, three of three, and one of four.

Sequences of three or more PPs are thus rare in English institutional discourse as revealed by the respective medians (median for sequences of three PPs = 1; median for sequences of 4 PPs= 0). In addition, sequences of four PPs are practically non-existent (in 85% of the data we find none). The median number of four or more juxtaposed PPs in the corpus is zero.

TABLE 10
SEQUENCES OF PPs PER 1,000 WORDS IN ENGLISH DISCOURSE

Frequencies Descriptive Statistics n=20 texts				
	Sequences with 2 PPs	Sequences with 3 PPs	Sequences with 4 or + PPs	TOTAL NUMBER OF JUXTAPOSED PPs IN THE CORPUS
Median	10	1	0	13
Min.	0	0	0	0
Max.	23	6	6	29
Range	23	6	6	29
Count per 1,000 words for the corpus as a whole				
	13	3	1	17

*NOTE: Totals may not add up due to rounding.

A look at the minimum and maximum number of occurrences of juxtaposed prepositional phrases per 1,000 words in the English data further reinforces that English writers rarely use prepositional phrases to develop discourse. The ranges are short.

Adverbial expressions of time and place

The figures for the use of locative and temporal adverbial phrases in the English corpus appear in Table 11. In the corpus as a whole, this type of attribute occurs 37 times in each thousand words. This number is similar to the measure of central tendency (median= 38; range=48). In addition, half of the texts present from 27 to 44 locatives and time adverbials per thousand words (50%). To determine the intervals in Table 11, I checked how the data was spreading from the median in terms of frequency of use. Then, I separated it in two extremes that were less frequently used and a center that occurred with a somewhat higher frequency.

TABLE 11

LOCATIVE/TIME ADVERBIALS PER THOUSAND WORDS IN ENGLISH TEXTS

Descriptive Statistics n=20 texts		ADVERBIALS	#	%
			Texts	
Median	38	17-26	4	20
Max. value	65	27-44	10	50
Min. value	17	46-65	6	30
Range	48			
TOTAL NUMBER OF OCCURRENCES:1000		37		

Independent sentences/clauses

In Table 12, I examine the occurrence of conceptually autonomous action chains or independent clausal constructions. The total number of occurrences of independent

constructions in the corpus is 53:1000. Descriptive statistics indicate that the median number of independent sentences/clauses is 50 occurrences (range=32). As we would expect, their distribution is uneven across texts (max=70; min=38) since texts vary in length and content. Most texts contribute one distinct number of independent clauses to the distribution.

TABLE 12
INDEPENDENT CLAUSES PER THOUSAND WORDS IN ENGLISH TEXTS
N=20 texts

Descriptive Statistics		Independent clauses	# texts	%
Median	50	38-44	5	25
Max. value	70	46	3	15
Min. value	38	48-49	2	10
Range	32	50	2	10
		53-61	4	20
		62	3	15
		70	1	5
TOTAL NUMBER OF OCCURENCES FOR THE CORPUS:1000		53		

* NOTE: Totals may not add up due to rounding.

In 55% of the cases (n=11) one could find from 46 to 61 independent clauses with a slightly greater chance of finding 46 (15%) or 62 (15%), and 50 (10%), in that order.

Summary

Table 13 summarizes the frequency counts and counts expressed per 1,000 words for the linguistic attributes in the English data, the type/token ratio and the median length of sentences and attention units. Results revealed that clause construction in the English sample is characterized by attention units of nine words; sentences of 20 words; and 53 independent clauses in each 1000 words. Furthermore, the sample typically conveys

information by means of the lexicon as reflected by the number of nominalizations and content words, and by the type-token ratio. The data not only employs many different words yielding a high lexical variety but more than half of those words (=61%) are content words, reflecting a precise word choice and high informational density. With regards to clause construction, nominalizations appear as the principal means whereby writers increase the size of attention units and package information. Compared to prepositional phrases, there are almost four times as many nominalizations (61:1000) as two or more juxtaposed prepositional phrases (17:1000) in the English corpus. Finally, there is low linguistic indication of involvement with concrete reality. Only 37 words in each thousand were locative or temporal adverbial phrases.

TABLE 13
EVENTS ASSOCIATED WITH ENGLISH TEXTS
number of words = 6398

Category	Median	Ratio
Sentence length	20	-
Attention unit length	9	-
Type-token		0.50
	Frequency count	Normalized Frequency
Prepositional Phrases (2 or +)	101	17
Nominalizations	394	61
Content words	3919	611
Independent sentences/clauses	340	53
Place/time adverbial phrases	238	37

*NOTE: Total counts/percentages in the data may not add up due to rounding.

Qualitative analysis of these attributes in the context of discourse shall enlarge our understanding of events associated with the “communicative text” category in English

institutional discourse. In the next sub-heading, I will examine how English combines events that are typically associated with it according to perceptual, cognitive, and pragmatic factors.

Explaining the COMMUNICATIVE English text: How do linguistic attributes cue comprehension in English?

For Cognitive Grammar (Langacker, 1991; see Chapter III for a review of the cognitive paradigm), any complex expression requires that readers establish correspondences between sub-parts and superimpose them to produce a final composite structure. Texts are no exception. This qualitative analysis will examine how the selected linguistic attributes, which are sub-parts of the PATH schema of a text, establish correspondences with sub-parts of other components of the COMMUNICATIVE TEXT image-schema. These other sub-parts, include cognitive and pragmatic factors.

The text we will examine comes from Shell's *The Test of Tomorrow* (1993; p. 3). It aims at informing readers about Shell's commitment to solve problems caused to the environment due to past human activities:

Past activities

Over the last decade, as understanding and aspirations with regard to protecting the environment evolved, society has become increasingly aware of the impacts of past human activity.

Shell companies recognise their responsibilities for their past activities and, where appropriate, act accordingly in compliance with prevailing legislation as a starting point. Experience has shown that the process to reach a well-balanced solution is often long and complex. Decisions—in general in consultation with the public authorities—have to be made on the most cost-effective means to achieve such a solution, frequently necessitating special research. Such a process may comprise a detailed inventory and investigation of the past activities and the environment in which they were conducted. This results in the setting of priorities and a plan based on an assessment of the environmental hazards involved, the significance of the impact and possible off-site migration of the pollutant.

In almost all instances where Shell companies undertook to act in respect of their past activities, this required considerable time, effort and funds. More than once the lesson learned was that cure is more expensive than prevention.

How does the lexicon cue comprehension in English? If we examine how the English lexicon has been used in this text, we see that 484:1000 words are content types while 591:1000 are content tokens. The other 392:1000 words encode grammatical functions in separate tokens. Out of the function words, only 102:000 are function types. In other words, many more types are content words and a few are function types.

The inventory of function types English uses is limited: *a, and, as, for, in, of, on, over, such, than, that, the, their, they, this, to, where, which, with*. There is considerable repetition of these same words (35%; 297:1000). Compared to content words (13%; 75:000), there is much more repetition of function words than of content words. English seems to use lexical density to cue meaning.

The content words which are repeated (*past* and *activity* with five mentions; *environment* with three mentions; *Shell* and *companies* with two mentions; *solution* and *process* with two mentions; *impact* with two mentions) reflect the topic of discourse: Shell companies' attitude toward the impacts caused by their past activities on the environment. When they are not repeated to cue back reference, they are implicated by the situational context. That is the case of the expressions in italics and bold:

Decisions [*about how to solve problems caused by past activities of Shell companies*]

Process [*of reaching a solution about how to solve problems caused by past activities of Shell companies through special research*]

More than once the lesson learned [*by Shell companies*] was that cure is more expensive than prevention.

The elided content tokens act as covert cohesive ties implicated by the situational context. Though they are not explicit, they are points of contextual reference. As points of contextual reference, they are reader conscious or semi-active information and require a relatively moderate amount of processing effort.

Reflecting the high number of content types and the relatively low number of repetition of content tokens, many of the content types are collocated to cue meaning. For example, *well-balanced solution with cost-effective means*; then *well-balanced solution with special research, process, inventory, investigation, setting priorities, plan, and assessment*. Collocations are also reader conscious or semi-active information. According to the connectionist approach to discourse processing, whenever we activate one node in a semantic field, other related nodes are immediately energized and made available for processing if necessary.

Verbs also play a role among the content types. Except for *to become*, *to act* and the auxiliaries *to have* and *to be* which have been repeated at least once each, verbs contributed to lexical density exhibiting a somewhat varied inventory. They extend discourse through the nominal forms—infinitives, gerunds, and participles—*to reach*, *to achieve*, *necessitating*, *based*, *involved*, *to act*—or by establishing contrast between past and present processes:

as understanding and aspirations with regard to protecting the environment **evolved**
 Shell companies **recognise** their responsibilities and where possible **act** accordingly...
 The process...**is** often long and complex.
 ...in which they **were conducted**.
 This **results** in the setting of priorities
 ...where Shell companies **undertook** ...
 This **required** considerable time...
 ...the lesson learned **was** that cure **is** more expensive than prevention.

Most of all, the temporal contrast verbs establish between past and present time gives prominence to Shell's commitment to the environment. All verbs that relate to past activities appear in the past and those that relate to the topic of discourse and the image they want to sell of Shell Companies appear in the present.

Another observation is that many more content types are nouns (220:1000) immediately followed by verb types (129:1000) and adjective types (91:1000). Adverbial types are scarcely used (48:1000). This observation seems consistent with the fact that English is developing the topic through lexical collocation and contrast of temporal processes.

With regard to function tokens, the definite article *the* (75:1000) and the conjoiner *and* (48:1000) were the most frequently used attributes. If we reason in terms of grammatical classes, the most used function tokens were prepositions (118:1000), particularly *of* and *in*, the definite article *the* (75:1000), and pronouns (59:1000). Among the pronouns, we have relative pronouns (27:1000), possessives (16:1000), deitics (11:1000) and personal pronouns (5:1000). These function tokens have ecological validity because they are frequently profiled. The more frequently a form is used in a language the more unmarked it becomes and the fewer cognitive operations it requires.

By examining any one excerpt, we understand why those function tokens were frequently used. Because English has few morphological resources, it repeats the

grammar words that cue givenness, that act as atemporal relations, that coordinate nominal expressions or that refer back to active or semi-active referents to develop the topic. That is the English way of connecting trajectors to landmarks:

Such a process may comprise a detailed inventory and investigation of the past activities and the environment in which they were conducted. This results in the setting of priorities and a plan based on an assessment of the environmental hazards involved, the significance of the impact and possible off-site migration of the pollutant.

In this excerpt, *such* refers back to the set of actions Shell companies are taking to solve problems caused to the environment by past activities. Then the indefinite article *a* as well as \emptyset article introduce some of the actions the Companies are taking—*a detailed inventory, \emptyset investigations of past activities, a plan based on an assessment, off-site migration*. We also see *and* conjoining, in three different mentions, the actions of the Companies; and the definite article introducing all expressions that collocate with the environmental theme (*the environment; the past activities; the impact of off-site migration of the pollutant*). Also, pronouns are typically acting as anaphors, referring back to nominals that construe the conceptual structure of this piece of discourse and that, as content words, have been repeated to cue meaning: *which (environment); they (past activities); this (such a process; the set of actions to solve problems of the past)*.

The examination of English texts also reveals that morphology, contrary to function words, does very little for readers in terms of tying concepts together and cueing meaning. In the piece above, for example, in only three cases do we see inflection for plural. In the three instances, the plural marker does not enter in correspondence with any other concept in the piece. Rather, it appears in profile determinant structures that

enter in head-complement relations: *investigation of the past activities; setting of priorities; assessment of the environmental hazards*. In addition, there is one inflection for third person singular that ties back to a nominal referent: *This (such a process) results...*

In the case of verbs or temporal processes, English draws on different content tokens to convey notions of aspect and mood:

society **has become** increasingly aware of the impacts of past human activities
 Experience **has shown** that the process ...
 Decisions **have to be made** on the most cost-effective means...
 Such a process **may comprise**

The need to draw on different content tokens to convey grammatical notions increases lexical density.

The number of prepositions and the repetition of the prepositions *in* and *of* relates to topic development and is consistent with the high number of nouns this piece of discourse presents. In most cases, prepositions appear in head-complement relations:

As understanding and aspirations **with** regard to protecting the environment evolved,
 Shell companies recognise their responsibilities **for** their past activities
 Such a process may comprise a detailed inventory and investigation **of** the past activities
 and the environment **in** which they were conducted
 the significance **of** the impact and possible off-site migration **of** the pollutant.
 Decisions in general in consultation **with** the public authorities
 Such a process may comprise a detailed inventory and investigation **of** the past activities
 This results in the setting **of** priorities and a plan based on an assessment **of** the environmental...
 the significance **of** the impact and possible off-site migration **of** the pollutant
 Shell companies undertook to act in respect **of** their past activities

In other words, this example illustrates how prepositions typically introduce entities that construe the topic:

With regard to	<i>protecting the environment</i>
of	<i>past human activity</i>
for	<i>their past activities</i>
with	<i>prevailing legislation</i>
with	<i>the public authorities</i>
of	<i>the past activities and the environment</i>
of	<i>priorities and a plan</i>
of	<i>the environmental hazards</i>
of	<i>the impact</i>
of	<i>the pollutant</i>
of	<i>their past activities</i>

Four mentions of the *past activities* theme and four mentions of the *environmental* theme appear as e-sites. That is, they are profile determinant not only in relation to their heads, but also in relation to discourse itself.

English readers activate the autonomous entities (e.g. *understanding* and *aspirations, responsibilities, significance* etc.); then they activate the dependent entities in which the e-sites appear (e.g. *with regard to protecting the environment, for their past activities, of the impact*) and superimpose it on the base. The conceptualization of the composite structure requires no further mental re-arrangement. It follows the interpretive ordering common to the categorization of any basic category: first we conceptualize the more inclusive member and then the more specific.

As to the preposition *in*, its repetition ties into the topic development as well. It appears in adverbials that modify the temporal processes and show how Shell companies act and make decisions to be environmentally friendly:

Act accordingly *in* compliance ...
 Decisions—in general *in* consultation
 To act *in* respect...

In brief, the qualitative analysis of these excerpts confirms and explains findings of the quantitative analysis for lexicon organization. English discourse uses lexical density and variety to cue meaning and give continuity to the discourse topic. However, what the quantitative analysis fails to show is that the most frequently used attributes are content types, specially nouns. They are the main lexical ties English texts use to weave “texture” as the examples illustrate. That is, English institutional expository discourse typically favors collocations of distinct words to cue meaning. The analysis has also shown that because English rarely combines different grammatical information in one single form, it needs to use different lexical items to cue notions such as mood and aspect. That increases the number of content tokens, and ultimately, lexical variety. Though the text in the example carried few function types, those that were present were repeated with high frequency. In other words, English seems to repeat function tokens to establish relations among content tokens and construe the topic: for example, the definite article *the*, conjoiners (=and) and prepositions (=of and *in*). Most prepositions introduce dependent entities whose e-sites are profile determinant in relation to the discourse topic.

How does syntax cue meaning in English? To begin with, sequences of PPs seem to be minor factors, as the example below illustrates:

Decisions **in** consultation **with** the public authorities
 the significance **of** the impact and possible off-site migration **of** the pollutant

In this case, there are two sequences of two juxtaposed PPs. This piece of discourse does not seem to favor sequences of PPs as a device to lengthen its sentences. However, PPs in combination with adverbial clauses do seem to collude to ground the independent constructions that foreground the main idea:

Over the last decade,

as understanding and aspirations

↓
with regard to **protecting the environment** evolved,

SOCIETY HAS BECOME INCREASINGLY AWARE

↓
of the impacts

↓
of past human activity.

Another aspect of syntax that the study undertook to analyze relates to attention units as a function of the number of words between punctuation marks. Below I list the attention units in the text:

PARAGRAPH I

- 1 Over the last decade, (4 words)
- 2 as understanding and aspirations with regard to **protecting the environment** evolved, (11)
- 3 society has become increasingly aware of the **impacts of past human activity**. (12)

PARAGRAPH II

- 4 Shell companies recognise their responsibilities for their **past activities** and where appropriate, (12)
- 5 act accordingly in compliance with prevailing legislation as a starting point. (11)
- 6 Experience has shown that **the process to reach a well-balanced solution** is often long and complex(16)
- 7 Decisions— (1)
- 8 in general in consultation with the public authorities— (8)
- 9 have to be made on the most **cost-effective means to achieve such a solution**, (14)
- 10 frequently necessitating special research. (4)
- 11 **Such a process** may comprise a detailed inventory and investigation of **the past activities and the environment** in which they were conducted. (22)
- 12 This results in **the setting of priorities and a plan** based on an assessment of **the environmental hazards** involved, (19)
- 13 the significance of **the impact** and possible **off-site migration of the pollutant**. (12)

PARAGRAPH III

14 In almost all instances where Shell companies undertook to act in respect of **their past activities**, (16)

15 this required considerable time, (4)

16 effort and funds. (3)

17 More than once the lesson learned was that cure is more expensive than **prevention**. (14)

Out of 17 units, 65% are lengthier than 10 words. A clause analysis of the units indicate that 17% of them constitute units with three clauses (attention units # 6, #12, and # 17); 41% constitute two-clause units (units #4, # 9, # 11, and # 14); and 41% clause-like units #2, # 3, #5, #8, # 10, # 13, and #15). Few attention units correspond to constituents (# 1, # 7, and # 16). There are units with up to 22 words constituting a single sentence (attention unit # 11) as well as units with 19 (#12) and 16 units (# 6 and #15).

In other words, generally, attention units in English do not help readers to parse. There are seven units (38:1000) that have more than one clause. When readers face this text and others in the sample, they need to allocate attention resources to parse it in clauses or units that can be under the focus of attention.

Apparently, to compensate for length of attention units, most sentences in the text have an average of 23 words. That is, the text has a moderately-high number of sentences with a moderately- low number of words. As we have seen, English writers use the situational context as well as the situational model to implicate content tokens that relate to the topic, relying on contextual effects.

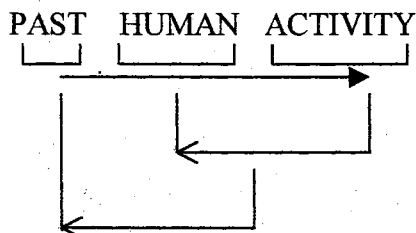
In addition, most mentions of the main theme appear in long attention units (see the words in bold in the previous example) that carry the autonomous entities in final constituent position. This means that many relational configurations of entities, though

activated serially, require mental rearrangement before the construal of a final composite structure. For example, in

past human activity

the landmark *past human* appears first, constituting the unprofiled part of the base that the head *activity* elaborates. However, until readers reach the head *activity*, they have no idea about the structure of this autonomous entity that is moving across discourse. While the landmark is stationary cueing ground, the trajector is profiled at the highest level, cueing figure. Actually, readers activate *past*, then *human* and finally they activate the autonomous entity *activity* that is profile determinant and elaborates the dependent entities *past human*. The interpretive ordering inverts the linear activation of entities. Readers need to re-arrange the forms to construe the final composite *past human activity*.

That takes time and effort:



Every head-modifier sequence in English requires rearrangements like the one illustrated above. Therefore, they demand attention resources. Because they are characteristic of the structure of English, they occur highly frequently. In this text, for example, we have

Last decade	Well-balanced solution
Past human activity	Public authorities
Past activities	Special research
Prevailing legislation	Detailed inventory and investigation
Starting point	Environmental hazards
Off-site migration	Past activities

Figure 13 further illustrates the need for mental re-arrangement in English head-modifier sequences. In it, I diagram the sequence *environmental impact studies* using Langacker's (1991) descriptive tools. To construe the final conceptual structure (represented by the heavy-line rectangle), users reverse (see the arrows) the order the entities appear.

Similar to the results of the quantitative analysis, qualitative analysis indicates that English uses independent constructions quite frequently. In this specific example, there are 48:1000:

1. society has become increasingly aware of the impacts of past human activity.
2. Shell companies recognise their responsibilities for their past activities and where appropriate,
3. [Shell Companies] act accordingly in compliance with prevailing legislation as a starting point.
4. Experience has shown...
5. Decisions have to be made on the most cost-effective means
6. Such a process may comprise a detailed inventory and investigation of the past activities ...
7. This results in the setting of priorities and a plan
8. this required considerable time, effort and funds.
9. More than once the lesson learned was

Independent constructions appear as figures or participants that are in motion to construe the discourse schematic-image. Along their path, dependent constructions, mainly adverbial phrases and non-finite clauses, provide the scenery and the details of the action:

1. as understanding and aspirations with regard to protecting the environment evolved
2. that the process
3. to reach a well-balanced solution is often long and complex
4. in general in consultation with the public authorities
5. to achieve such a solution
6. frequently necessitating special research

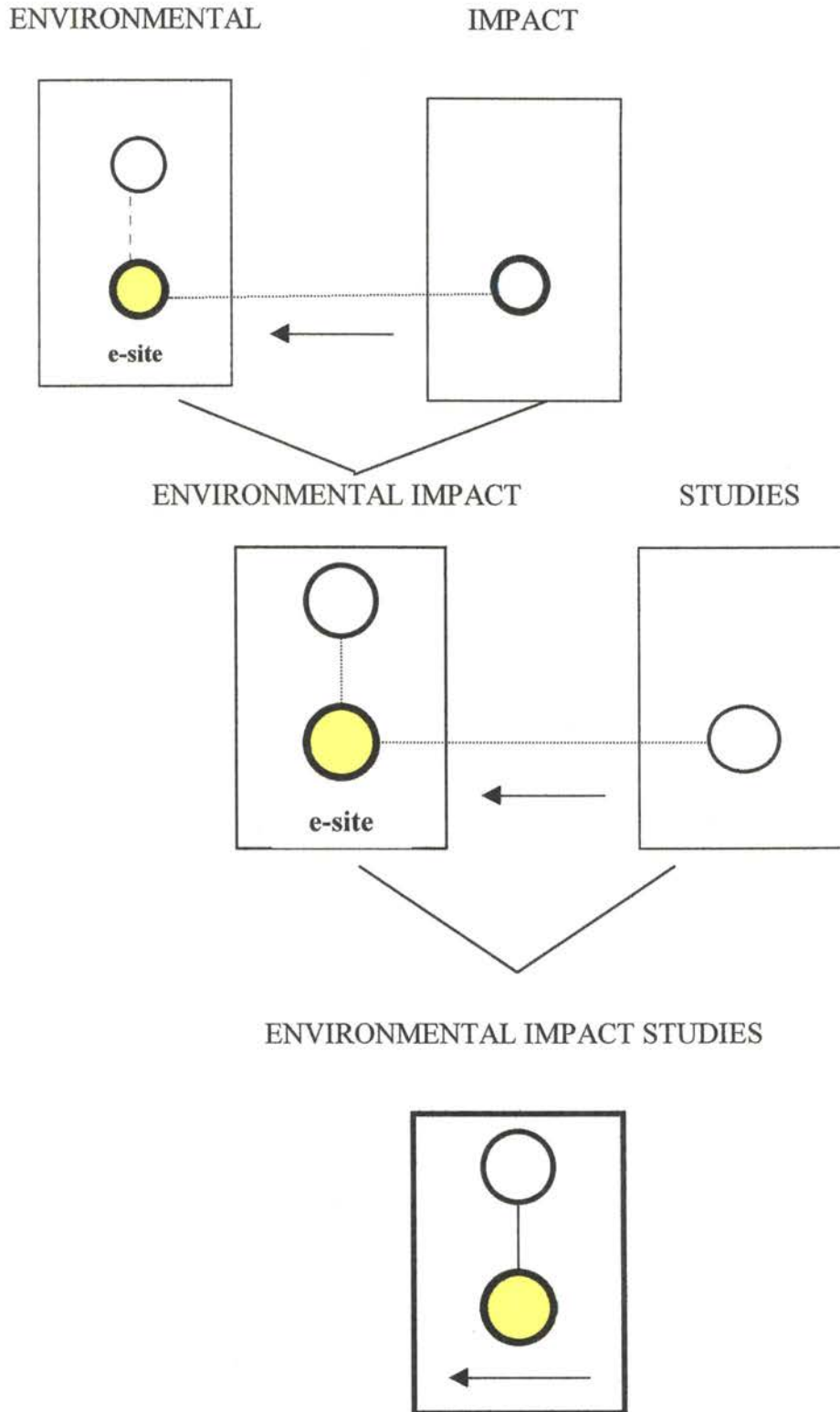
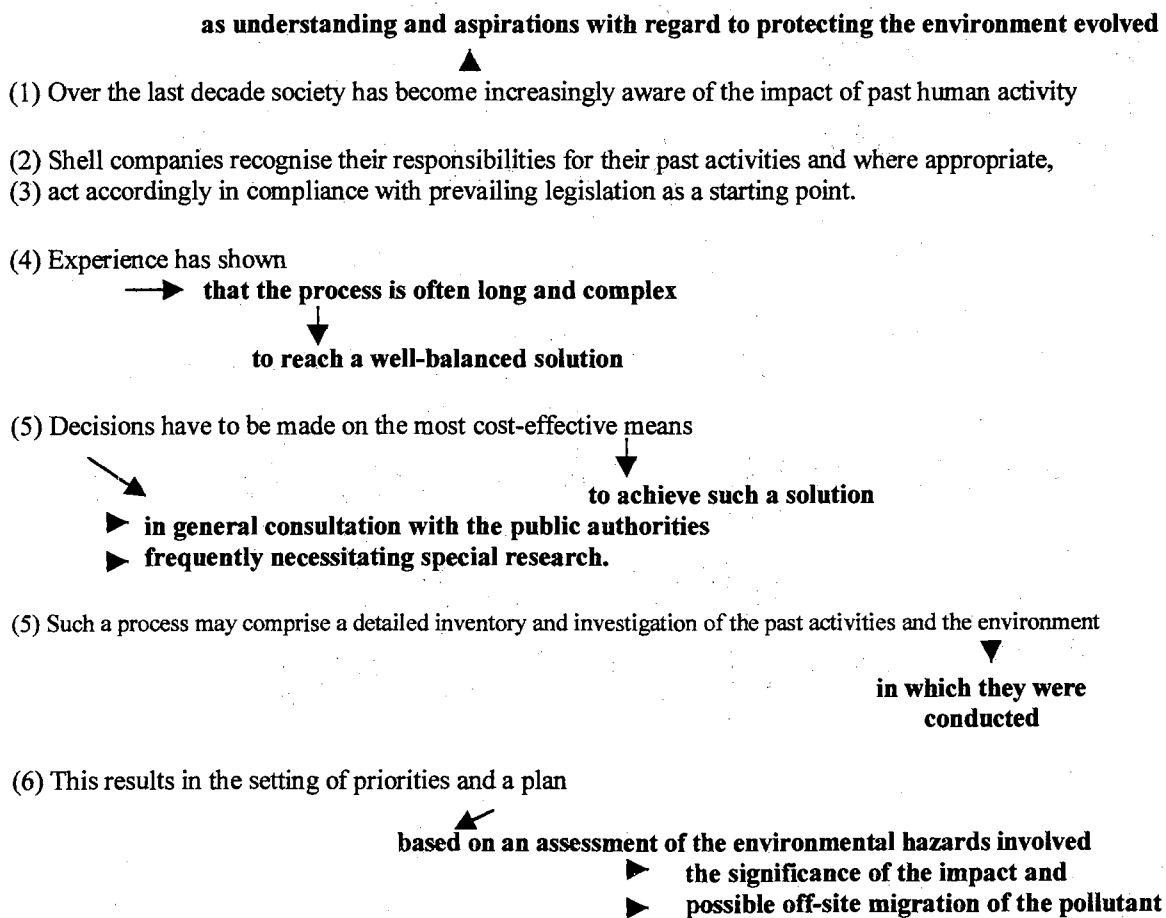
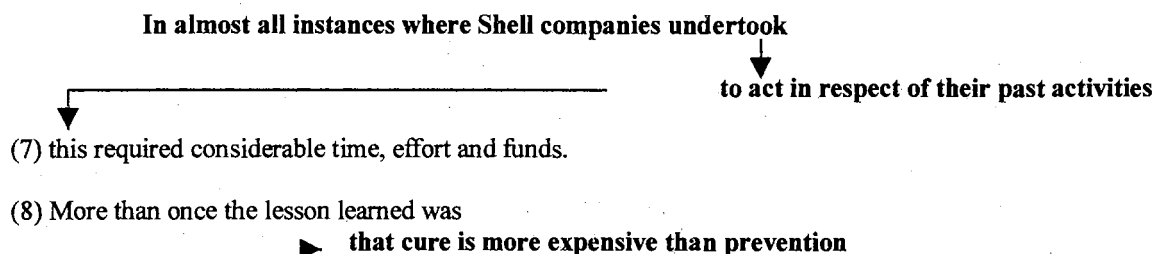


Figure 13. Conceptualizing syntagmatic relations in English.

7. in which they were conducted
8. plan based on
9. an assessment of the environmental hazards involved the significance of the impact and possible off-site migration of the pollutant
10. to act in respect of their past activities
11. that cure is more expensive than prevention

The example reveals that dependent constructions also appear with high frequency (59:1000), a detail quantitative analysis could not show. Actually, this English text draws slightly more on dependent than on independent constructions to develop its topic. Most dependent structures are non-finite infinitive *-ed* or *-ing* clauses with a greater tendency to be infinitive non-finite clauses (16:1000). A diagrammatic representation of the clauses would look like:





The dependent constructions in the diagram appear in bold. They not only surround the independent constructions but also elaborate on them. The arrows indicate the interpretive order. For example, the clause *this required considerable time, effort and funds* is the trajector or figure of the clause *in almost all instances where Shell companies undertook*, the landmark. The e-site of this structure is another dependent construction that is profile determinant—*to act in respect of their past activities*. In another instance, the entity *This results in the setting of priorities and a plan* construe the figure on which the dependent construction *based on an assessment of the environmental hazards involved* elaborates. The latter entity, for its turn, defines the entities *priorities and plan* that appear in the independent construction. Likewise, the entities *as the environmental hazards involved, the significance of the impact, and possible off-site migration of the pollutant* define the entity *assessment*.

Furthermore, the diagrammatic structure of the sentences in this piece of discourse indicates that though right branching constructions such as non-finite and other dependent clauses appear with frequency, their number is normally balanced by the number of autonomous action chains. The dependent clauses are like satellites. Also, with the exception of the first, fifth, and seventh sentences, all the others may be conceptualized in the order their sub-components (=the clauses) appear.

Another attribute this study selected to analyze because of its processing function are nominalizations. In the English sample, nominalizations appear somewhat frequently, as the quantitative analysis indicates. According to the qualitative examination of the texts, this seems also to be the case. In this particular example, we find 86:1000 nominalizations:

understanding and aspirations with regard to protecting the **environment**
 past human activity
compliance with prevailing **legislation** ...
experience has shown that...
 well-balanced **solution**
decisions have to be made...
 in **consultation** with...
 to achieve such a **solution**
investigation of the past activities and the **environment**...
assessment of the environmental hazards involved
significance of the impact
off-site migration of the pollutant
 cure is more expensive than **prevention**

They appear either as sources of energy in the action chains (see the nominalizations in the first sentence for example) or enter head-complement (e.g. *a plan based on an assessment of the environmental hazards involved*) and head-modifier (e.g. *prevailing legislation; off-site migration*) relations as figures. When in head-complement relations, their complements, as in other head-complement relations that we have examined, are profile determinants (e-sites) and connect them to the environmental theme. The fact that their landmarks are profile determinants reflect the relevance of nominalizations to topic continuity. In other instances, the complements themselves are nominalizations (e.g. **compliance** with prevailing **legislation**).

As the review of the literature has indicated, nominalizations involve reification. That is, every time a nominalization appears, readers need to implicate a bounded region

(a nominal) in which coordinated states, conceived temporally, combine. For example, *assessment* and its verbal form *to assess* can both be used to describe the same event:

a plan based on an assessment of the environmental hazards involved
a plan that assessed the environmental hazards involved

Though the two options profile the same conceptual structure [ASSESS], they use different cognitive routes to get there. While *to assess* imposes a processual construal on the profiled event, *assessment* portrays it as an abstract region. Because nominalizations such as *assessment* lay components one on top of the other in their conceptual structure, they demand mental re-arrangement, and ultimately, attention resources.

In some instances, though, nominalizations yield contextual effects that lead to processing gains. That is the case of *environment* in the example we are analyzing. Text-external sources, such as the current ecological movement, and the frequency with which the environmental theme appears in the media cumulate to maximize the relevance of the nominalization. Had the writer chosen another form of expression, it would have been more costly to attain the communicative goal. Text-internal sources such as collocation networks with the theme *protecting the environment* also maximize the relevance of that nominalization and of nominalizations such as *solution*, *decisions*, and *prevention*.

In brief, the qualitative analysis of syntactic organization in English indicates that though English favors independent constructions, it also draws frequently on non-finite clauses to establish interconnectedness between entities in discourse. In addition, English seems to prefer non-finite clauses to juxtaposed PPs as a right-branching device to develop the topic. As opposed to non-finite clauses, juxtaposed PPs appeared in only a

few instances in the examples analyzed. Furthermore, attention units most of the time failed to help readers of the sample text in the parsing task. Half of the units had three or two-clauses each; the other half had clause-like units. The clause-like units tended to be longer than 10 words, carrying nominalizations within it. Nominalizations were used with some frequency, generally appearing as figures or as e-sites. Some of the nominalizations tied into the topic and yielded cognitive gains. While attention units were long, sentences tended to have at the most 23 words. Finally, qualitative analysis has shown that in many instances word order failed to allow conceptualizers to mentally arrange the discourse entities in the linear order they appear. Those instances required mental rearrangement and, therefore, attention resources.

How do locative and time adverbials cue interaction with reality? The only AdvP of time constitutes the very first entity to appear in the text—*Over the last decade*. Its position is strategic because it refers to the preceding time context that is common to all readers. It could easily have been an inferrable entity. However, to emphasize since when *society* (including the readers and the main actor, *Shell*) has been environment-oriented, the period of time is explicitly mentioned and appears in first position. That makes it contextually and perceptually salient, and maximally comprehensible. In that it is maximally comprehensible, the AdvP implicates that a process has been going on and that readers are taking part in it.

The Adv.P. contextual effects are reinforced by the adverbial clause *as understanding and aspirations with regards to the environment evolved*. The two expressions call on the audience's own experiences with environment-related topics through that decade and on what they have learned to aspire. That is, the expressions involve the reader in the process and imply that anything that happened before the

process had begun was true of *society*. *Shell* (a semi-active entity at this point of discourse) has simply been in line with people's understanding and aspirations with regard to the environment.

In fact, the themes (understood as the left-utmost element of each sentence) indicate not only the GOAL of this piece of discourse but also its ORIENTATION. The Adv.P. appears first. Then, *Shell Companies*, the main actor and source of energy, appears two times in sentence initial position as displaced information. In other instances, the major discourse participant is assumed to be current-given because of the magazine in which the text has been published (a Shell publication) and because of the publication's GOAL—to inform readers about Shell's activities. That information is part of the situational model. In addition, after its first explicit mention, *Shell* is implicated by maximization of contextual clues in different instances. Some of them are

(Shell companies) act accordingly.....
 Decisions (by Shell companies) have to be made
 This required considerable time, effort and fund (from Shell companies)
 More than once the lesson learned (by Shell companies) was that cure is...

The cognitive effects generated by contextual clues make next-mentions of the main actor not relevant and less efficient in terms of processing effort.

If we use clauses as the unit of analysis, we see that the AdvP as a theme also cues ALIGNMENT. The themes interactionally negotiate role relationships and constitute face-saving devices. Shell's role is to act in line with people's aspirations. In a way, the introduction reinforces the effects generated by the themes, appearing as a rhetorical

strategy of presentation or a staging device that marks Shell's solidarity with the audience's concerns.

Summary. The qualitative analysis of this text combined with findings of the quantitative research indicate that English uses lexical density, particularly by means of collocating content types and word order/syntax to maximize contextual clues and cue comprehension. Specifically, the qualitative analysis demonstrates that English repeats function tokens frequently, mainly prepositions and the definite article. Those words establish ties or trajector-landmark relations among content words in the text. Also, 23-word-sentences organized in attention units with 11 words on average, frequent non-finite and independent clauses are syntactic devices English uses to yield maximal cognitive effects. The attention units resemble clause-like units and each sentence shows on average two attention units. With regard to word order, the qualitative analysis indicates that though head-complement relations in English favor the conceptualization of component structures in the order they appear, head-modifier sequences force readers to rearrange components in a cognitive route that differs from the one in which the words appear. While juxtaposed PPs appear rarely, nominalizations appear with some frequency and most of them demand attention resources, forcing the reader to lengthen the cognitive route. Last, the qualitative analysis rarely found AdvPs of time and place. The only AdvP of time that did appear related to the GOAL, and ALIGNMENT of the discourse and signaled interaction with the audience's reality. The analysis also found other indications of interaction with reality as for example, the adverbial clause of time that immediately followed the AdvP, the use of words such as *society* including the reader and the displaced but reader-conscious ADDRESSER in the same situational model.

Brazilian Portuguese

To find out the events that are typically associated with BP texts, I counted the frequency of occurrence of the selected attributes in BP institutional discourse and normalized those figures. As in English, the Brazilian Portuguese COMMUNICATIVE TEXT category also activates knowledge of a series of events.

As we did for attributes associated with English discourse, next we will treat each event associated with BP institutional discourse in turn.

Sentence Length

Table 14 and 15 outline the distribution of sentences in Portuguese as a function of the number of running words. Table 14 reproduces the median frequency of words per sentence and the variability of the data from the central tendency (Median=24; range=47). The most typically used sentences have between 17 and 31 words. The range (=47) confirms the spread of the variation.

TABLE 14
SENTENCE LENGTH IN BRAZILIAN PORTUGUESE

n=241 sentences

Descriptive Statistics		Number of words	<i>f</i>	%
Median	24			
Max. value	55	08-16	51	17.4%
Min. value	08	17-31	129	53.4%
Range	47	32-42	41	18.6%
		43-55	20	10.1%

A closer look at the individual frequencies per sentence reveals that many more sentences group at the center of the distribution ($n=129$; 53.4%) and toward the upper end of the scale ($n=41$; 38.6%) than toward the lower extreme. Out of a total of 241 sentences in the BP data, 69 (28.7%) appear in the upper end of the range, having from 32 to 55 words and only 17.4% ($n=51$) of the sentences group at the lower end of the range. Those findings reflect a slightly negatively skewed distribution. Sentences in BP tend to be longer than 17 words.

To determine the intervals in Table 14, I looked at the frequency counts. Sentences that occurred moderately ($2 \leq f \leq 8$) appear first; sentences that occurred more than 9 times or with high frequency ($9 \leq f \leq 13$) appear second; sentences that also occurred moderately ($2 \leq f \leq 8$) appear third; and sentences that rarely occurred ($3 \leq f \leq 1$) appear last. Table 15 rank orders the most used sentence lengths in the sample and outlines the percentages and frequency of use as a function of the number of words. The most widely used sentence length in the BP data is 20 words per sentence ($f=13$; 5.4%), immediately followed by sentences with 21 ($f=12$; 5%), 19 ($f=11$; 4.6%), and 18 ($f=11$; 4.6%).

The results in Table 15 reinforce the moderate tendency BP writers have to use sentences that cluster in length at around the median number of words (=24) or if not, spread out toward the upper end of the range. In BP we see sentences with 31 words appearing 10 times in the data as well as sentences with 40 words appearing 5 times. Even sentences with 49, 50, 52 and 53 words show more than once. Apparently sentence length in BP does not seem to be a constraint writers face in producing written language.

TABLE 15
 MOST USED SENTENCE LENGTH IN BP
 number of sentences = 241

RANK ORDERED

Rank	Length	<i>f</i>	%
1	20	13	5.4
2	21	12	5.0
2	19	11	4.6
3	18	11	4.6
3	24	10	4.1
3	31	10	4.1
4	23	9	3.7
4	17	9	3.7
4	27	9	3.7
5	13	8	3.3
5	22	8	3.3
6	15	7	2.9
7	32	7	2.9
7	33	7	2.9
8	25	6	2.5
8	26	6	2.5
8	28	6	2.5
8	30	6	2.5

Figure 14 illustrates the different sentence lengths in BP. The predominance of sentences with more than 17 words is particularly registered by the number of bars at the upper end of the scale and, specially, by the bar that represents sentences with more than 38 words. They constitute 16.2% of all sentences in the sample.

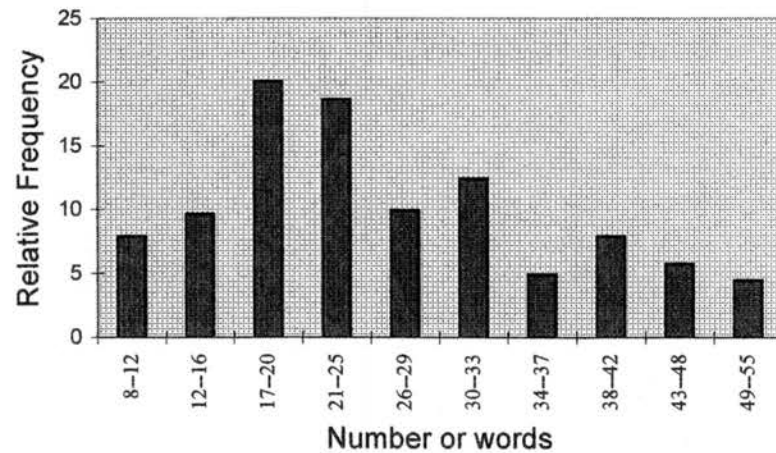


Figure 14. Sentence Length in Brazilian Portuguese

Brazilian Portuguese clearly organizes words in sentences that generally have from 17 to 33 words or that tend to be longer than 33 words.

Length of attention units

The distribution of words per attention units is represented in Table 16. While sentences in the BP data tend to be lengthier than 17 words, attention units tend to be shorter than 7 words. Out of 779 attention units in the data, 401 (51%) are from one to seven words long. The data is clearly positively skewed if you consider that the median is seven. In addition, if we eliminate units that were used with low frequency ($f < 18$; 16-32 words), we still remain with 729 (93.6%) units in the data. The range itself accounts for such tendency. The longest attention unit in BP has 32 words and the shortest has one word (range=31).

TABLE 16

LENGTH OF ATTENTION UNITS IN BRAZILIAN PORTUGUESE

n=779 attention units

Descriptive Statistics		Number of words	<i>f</i>	%
Median	7	1	44	5.6
Max. value	32	2	63	8.1
Min. value	1	3-7	294	37.7
Range	31	8-9	125	16.0
		10-16	203	25.9
		17-32	50	6.4

The intervals in Table 16 were determined according to the frequency with which an attention unit occurred. Attention units that have from 17 to 32 words, rarely occurred ($1 \leq f \leq 13$); attention units that have from 10 to 16 words occurred with low frequency ($39 \leq f \leq 18$); those that have from 3 to 9 words occurred with high frequency ($58 \leq f \leq 72$). This last interval was sub-divided into two (attention units with 3-7 and with 8-9 words) to better illustrate the variation. Attention units with eight words occurred 67 times and those with 9 occurred 58 times. Figure 15 depicts such distribution. The bars follow the intervals indicated in Table 16:

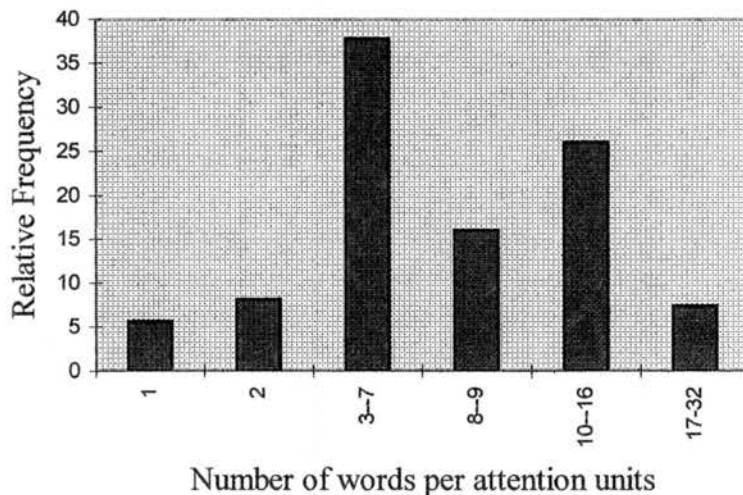


Figure 15. Relative frequency of attention units according to number of words.

These observations build robust evidence towards the typicality of attention units with up to 7 words. The bar that represents units with three to seven words stands out.

To further support the claim about the length of attention units in BP, Table 17 rank orders the most used attention unit lengths. It also shows each unit's relative frequency as a percent of the overall corpus of attention units ($n=779$). The most typically used attention unit is five words long ($f=72$; 9,2%), immediately followed by units with eight ($f=67$; 8,6%) words, and units with two ($f=63$; 8,1%) words. While units with five words occurred 72 times, the median occurred 58 times (7,4%), ranking at the fifth place. In addition, though units with eight and nine words appear in the second and fifth rank respectively, they are outnumbered in frequency by the units with two (third rank), four (fourth rank), three (sixth rank), and six (seventh rank) words.

TABLE 17
 MOST USED ATTENTION UNITS IN BRAZILIAN PORTUGUESE ACCORDING TO LENGTH
 number of units = 779
 RANK ORDERED

Rank	Length	<i>f</i>	%
1	5	72	9.2
2	8	67	8.6
3	2	63	8.1
4	4	62	8.0
5	7	58	7.4
5	9	58	7.4
6	3	52	6.7
7	6	50	6.4
8	1	44	5.6
9	10	39	5.0
9	12	39	5.0
10	11	37	4.7

The relevance of these results rests on the consistency with which the units are kept within the bounds of the median, or seven words long (51.5%), suggesting a typical length for attention units in BP.

Type-token ratio

Out of 6371 tokens in the Brazilian Portuguese data, 2962 were unique or different types, yielding a moderate type-token ratio of 0.49. This ratio indicates that more than half of the words (54%) in the BP corpus are repeated words. A look at the distribution of ratios reveals that individual texts vary somewhat in the way they use types and tokens (min= 0.36; max=0.67; range=0.31). As Table 18 shows, in half of the texts (n=10) we find type-token ratios of 0.36% to 0.48% and in the other half we find type-token ratios of 0.50% to 0.67.

TABLE 18
DISTRIBUTION OF THE TYPE-TOKEN RATIO ACROSS BP TEXTS
n=20 texts

Descriptive Statistics		type-token	<i>f</i>	%
Median	0.49	0.36 - 0.48	10	50
Max. value	0.67	0.50 - 0.67	10	50
Min. value	0.36			
Range	0.31			
TOTAL OCCURRENCES:1000		465		

*NOTE: Totals may not add up due to rounding

Those numbers confirm that Brazilian Portuguese texts tend to have a moderate number of different words.

Content words

Though lexical variety in BP yielded a moderate ratio, lexical density is relatively high. As Table 19 outlines, more than half of the words in the sample (553:1000) are content words. Actually, in 13 (65%) texts the number of content words is higher than 550:1000. While six of them (30%) show a proportion of 550:1000, four texts (20%) show 570:1000, two (10%) show 560:1000, and one (5%) show 590:1000. The tendency then is toward the upper end of the range (min=480:1000; max=590:1000; range=110:1000).

Specifically, Table 19 lists the frequencies—expressed per 1,000 words—for content words and gives the absolute and relative frequencies for each of the 20 texts. Given that we have moderate lexical variety in BP, it may be that writers repeated some content words. In the qualitative analysis we shall go back to this issue.

TABLE 19
 NUMBER OF CONTENT WORDS PER 1,000 WORDS IN BP
 n=20 texts

Descriptive Statistics		content words	<i>f</i>	%
Median	550	480	2	10
Max. value	590	510	1	5
Min. value	480	520	1	5
Range	110	530	2	10
		540	1	5
		550	6	30
		560	2	10
		570	4	20
		590	1	5
TOTAL OCCURRENCES:1000		553		

*NOTE: Totals may not add up due to rounding

The pie chart in Figure 16 illustrates the results discussed above.

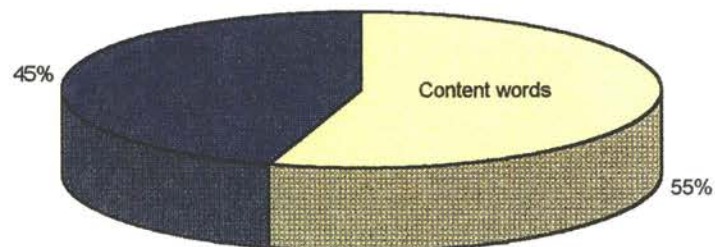


Figure 16. Typicality of content words in institutional BP discourse.

Nouns, verbs, adjectives, and adverbs constitute the majority of the words in the BP corpus (55%) and occur with high frequency. However, function words also occur with some frequency (45%). Thus, with regard to the lexicon, BP seems to combine grammatical and lexical information to cue meaning with a slight tendency to rely more on the lexicon.

Nominalizations

Table 20 outlines the distribution of nominalizations in the BP corpus (min=34:1000; max=113:1000) and provides a breakdown of their frequencies of occurrence per 1,000 words across texts in the data. By examining the distribution, we see, similarly to the English corpus, that there is no single number that stands out as the most typical, but a variety of numbers that evenly spread out from the median (Median=61; range=79). The only number that appears twice is 47:1000. A high number of cases (45%; n=9) concentrate at the center of the distribution, displaying from 45 to 62 nominalizations per thousand.

TABLE 20
NUMBER OF NOMINALIZATIONS PER 1,000 WORDS IN BP
n=20 texts

Descriptive Statistics		Nominalizations	f	%
Median	61	34-38	2	10
Max. value	113	45-48	4	20
Min. value	34	54-62	5	25
Range	79	71-77	3	15
		81-86	3	15
		101-113	3	15
TOTAL OCCURRENCES:1000			65	

*NOTE: Totals may not add up due to rounding

The bar chart in Figure 17 illustrates the frequency of nominalizations per thousand words in BP texts and the frequency with which values that range from 62 to 45 nominalizations per thousand words occur. Each bar represents an interval as indicated in Table 20. Namely, from left to right, bar number one provides the percentage of cases that display from 34 to 38:1000 nominalizations.

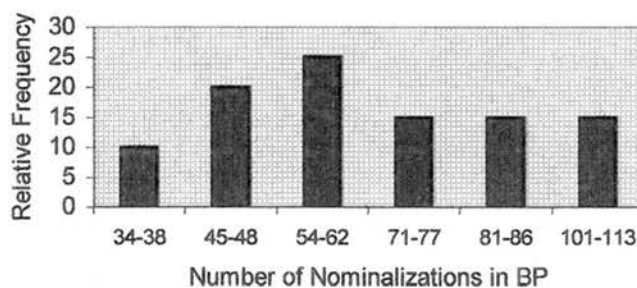


Figure 17. Occurrences of Nominalizations in BP

Thus, BP tends to use nominalizations with high frequency. According to the chart, this frequency varies from the center toward the upper end of the range (45-113 nominalizations).

Sequences of Prepositional phrases

The total number of juxtaposed prepositional phrases in the BP data is 35:1000. Out of that total, 22:1000 represent sequences with two juxtaposed PPs; 9:1000 represent sequences with three juxtaposed PPs; and 4:1000 represent sequences with four or more

juxtaposed PPs. Table 21 reflects those findings. It also outlines the distribution of PPs in individual BP texts, listing the median frequency of occurrence for each category investigated and the range. The study of the distribution reveals a median of 31 juxtaposed PPs for the twenty texts investigated (Median=31; range=54). The median for sequences with two PPs is 21 (range=49); sequences with three PPs is 8 (range=28); and sequences with four or more PPs is 3 (range=11).

TABLE 21
SEQUENCES OF PPs PER 1,000 WORDS IN BP

Frequencies Descriptive Statistics n=20 texts				
	Sequences with 2	Sequences with 3	Sequences with 4 or +	TOTAL Juxtaposed PPs
Min.	5	0	0	14
Max.	54	28	11	68
Range	49	28	11	54
Median	21	8	3	31
Count per 1,000 for the corpus as a whole				
	22	9	4	35

*NOTE: Totals may not add up due to rounding.

Though sequences with three and four PPs occur with low frequency, there is wide variation in the corpus (min=0; max=28 for sequences of three; min=0; max=11 for sequences of four). Furthermore, among sequences with four or more PPs we find some with four, six or even seven prepositional phrases as the qualitative analysis will show.

The range (min=34:1000; max= 68:1000) reinforces BP's tendency to rely on juxtaposed PPs to develop the topic. Furthermore, in only six cases there were not sequences of four or more PPs.

Table 22 further lists the number of observations of juxtaposed PPs in each individual text and their respective relative frequencies (percentages).

TABLE 22
TYPICALITY OF PREPOSITIONAL PHRASE SEQUENCES PER 1000 WORDS IN BP
N=20 texts

Number of Sequences	# texts	%
14	1	5
25-29	6	30
30-34	7	35
40-44	3	15
50	2	10
68	1	5

*NOTE: Totals may not add up due to rounding.

Most texts (35%) show a range of 30-34 sequences of PPs and other 30% show from 25 to 29 sequences of PPs. This indicates that 65% of the texts have at least 25 sequences or two or more juxtaposed PPs. We can then conclude that BP has a moderate tendency to cluster sequences of two or more PPs (35:1000) and a weak tendency to cluster sequences of three and four PPs.

Adverbial expressions of time and place

The figures for the use of locative and temporal adverbial phrases in the BP data appear in Table 23. The count of locative and time adverbials per 1,000 words for the corpus is 37. A study of the distribution across individual texts reveals a median frequency of use for these adverbials of 38 (range=64).

TABLE 23

LOCATIVE/TIME ADVERBIALS PER 1,000 WORDS IN BP

Descriptive Statistics n=20 texts		ADVERBIALS	# texts	%
Median	30	0	1	5
Max. value	63	23-30	10	50
Min. value	0	36-41	4	20
Range	64	47-63	5	25
Total	37			
occurrences:1000				

*NOTE: Totals may not add up due to rounding.

In addition, Table 23 also shows that in 50% of the texts the number of adverbials per thousand words varies from 23 to 30:1000; in 20% of the texts it varies from 36-41:1000; and in another 25% it varies from 47 to 63:1000. That is, in most texts the number of time and place adverbials is greater than 30:1000.

Independent sentences/clauses

Conceptually autonomous action chains appeared 40 times in each 1000 words in the Brazilian Portuguese corpus. The individual number of observations per text, the absolute and relative frequencies, as well as descriptive statistics are given in Table 24.

Though there is variation in the data (median=36; min=25; max=54; range=29), 45% of the texts show from 31 to 39 independent clauses/sentences per thousand words. The most frequent number of independent clauses is 36 (n=3; 15%), the median, immediately followed by 32 and 27 (n=2; 10%).

TABLE 24

INDEPENDENT CLAUSES PER THOUSAND WORDS IN BP

Descriptive Statistics n=20 texts		Independent clauses	# texts	%
Median	36	25-27	3	15
Max. Value	54	31-39	9	45
Min. Value	25	40-48	5	25
Range	29	50-54	3	15
TOTAL OCCURRENCES:1000		40		

*NOTE: Totals may not add up due to rounding.

We may conclude that BP institutional discourse typically shows at least 31 independent sentences in each 1000 words.

Summary

Table 25 summarizes events of BP institutional discourse by displaying frequency counts and counts expressed per 1,000 words for the selected attributes, the type/token ratio and the median length of sentences and attention units. Syntactic construction in BP typically keeps attention units within the bounds of 7 words (Median=7; range=31) and sentences within the bounds of 24 words (Median=24; range=47). In addition, BP writers

use 35:1000 sequences of PPs, a right-branching device known to increase the size of action chains. The number of juxtaposed prepositional phrases (35:1000) apparently does not relate to the size of the attention units, but definitely may relate to the length of sentences. Likewise, BP frequently draws on nominalizations (65:1000), another device writers use to increase attention units. Sequences of prepositional phrases seem to stand out as devices writers use to elaborate nominal reference and pack information in BP. Conversely, BP show a moderate number of independent clauses/sentences (40:1000).

With regards to lexical specificity, the type-token ratio (0.46) points toward some repetition in BP. Less than half of the words in the sample are types (2962: 6371). Furthermore, the number of content words (553:1000) in the data indicates a moderate level of informational density (0.55). Writers choose words with some precision. Last, the number of adverbial expressions reflects low involvement with concrete reality (37:1000).

TABLE 25
EVENTS ASSOCIATED WITH PORTUGUESE TEXTS
number of words = 6371

Category	Median	Ratio
Sentence length	24	-
Attention unit length	7	-
Type-token		0.46
	Frequency count	Normalized Frequency
Prepositional Phrases (2 or +)	225	35
Nominalizations	413	65
Content words	3524	553
Independent sentences/clauses	255	40
Place/time adverbial phrases	237	37

*NOTE: Total counts/percentages in the data may not add up due to rounding.

As in English, qualitative analysis of these attributes in the context of discourse shall shed light on the events associated with the COMMUNICATIVE TEXT category in Brazilian Portuguese institutional discourse. In the next sub-heading, I will examine how BP combines events that are typically associated with it according to perceptual, cognitive, and pragmatic factors.

Explaining the COMMUNICATIVE BP text: How do linguistic attributes cue comprehension in BP?

Similar to our qualitative analysis of English texts, first we will address the organization of the lexicon, then syntactic matters, and close with how BP texts interact with reality. The analysis will follow the premises of cognitive grammar, according to which whenever we have a complex structure, we assemble its sub-parts into a final composite cognitive structure such as an image-schema.

How does the lexicon cue comprehension? A fast look at the BP texts suffices to indicate that BP encodes different grammatical functions in one token. Take the text *Esforço para preservar o futuro* (=Striving to safeguard our future) extracted from Environment, Quality, and Safety (Petrobras, 1994; please refer to Appendix C for a complete text and gloss):

Consciente de que meio ambiente, segurança industrial e qualidade
 Certain (s.) that environment safety (f.) industrial and quality

são interdependentes, a Petrobras reuniu
 are interdependent (pl.) the (f.s.) Petrobras brought together (3rd.p.s; p.perf.)

a coordenação destas áreas num único
 the coordination of these (f.pl.) areas (f.pl.) (under one) single (m.s.)

orgão, a Susema, ligado diretamente à
 organization (m.) the (f.s.) Susema, connected (m.s.) directly to the (s.f.)

Presidência, *procurando* *integrá-las* *à*
company chairman (s.f.) **trying** **to integrate them (f.pl.)** **to the (f.s.)**

missão principal da *Companhia: produzir, comercializar*
mission (f.s.) **main** **of the (f.s.)** **company (f.s.): to produce, commercialize**

e transportar petróleo e derivados para garantir
and distribute **crude oil and** **oil products (m.pl.)** **to guarantee**

o abastecimento do País.
the (m.s.) **supplying (N; m.s.)** **of the (m.s.)** **country (m.s.)**

In this first paragraph, as well as in the other two paragraphs, BP encodes different grammatical functions in a single token (see expressions in bold). Morphology is mainly responsible for that. In one form, BP can express person, number, tense, and mood (e.g. *reuniu*; third person singular, preterit perfect; indicative mood). In most cases expressions carry gender and number trappings. In addition, there are also expressions that have been gramaticized in the language such as *reuniu* (brought together), *num* (under one), *petróleo* (crude oil), and *derivados* (oil products). Also, BP has clitics that function as anaphors:

procurando *integrá-las*
trying **to integrate them (f.pl.=the three areas)**

In the above example, *-las* refers back to the three areas Petrobras has brought together under one division; namely, *-las* keeps *environmental protection, industrial safety and quality of services and goods* active. That happens in other instances:

Este esforço integrado, que pretende assegurar
This (m.) **effort (m.s.)** **integrated (m.s.)** **that** **intends** **to safeguard**

a integridade do homen, compatibilizar a
 the (f.s.) integrity (f.s.) of (m.s.) mankind (m.s.) reconcile the (f.s.)

produção... com o meio ambiente e
 production (f.s.) of goods with the (m.) environment (m.) and

garantir a qualidade de produtos e serviços,
 guarantee the (f.s.) quality (f.s.) of products and services

denomina-se *Processo Petrobras de Meio Ambiente, Qualidade e Segurança.*
 calls itself Process Petrobras of Environment, Quality and Safety.

The clitic *-se* attaches to the verb, making the subject indeterminate but implicated by contextual clues. That is, though who gave this effort a label is not explicitly mentioned, contextual clues implicate maximally that it was Petrobras. Also, the clitic *-se* makes the sentence passive. A cognitive translation of it would be: This process is called Agenda for Environmental Protection (by Petrobras Directory). According to the Relevance Principle, by implicating what has been already established in discourse—Petrobras is the main source of energy--we minimize the processing effort by drawing on an economic motivation.

Another instance of a clitic functioning as an anaphor and guiding cognitive operations is

Para melhor conduzi-los, a Petrobras....
 To better conduct **them** (m.pl; its ventures), Petrobras

The reference to the anaphor *-los* appears in the previous sentence:

com as comunidades afetadas por seus empreendimentos
 with the (f.pl.) communities affected (f.pl.) by its (Petrobras') ventures (m.pl.)

Because Brazilian Portuguese tends to accumulate information in a single form and grammaticize others, its amount of content and function tokens end up being lower than that of other languages that lack morphologic trappings or that fail to have one autonomous expression to convey a specific meaning. This tendency may be specially strong if you consider function tokens and morpho-syntactic processes that guide attention in discourse. We shall come back to this issue when we compare BP to English.

In the case of this text, I found 607:1000 content and 392:1000 function tokens. Out of these 462:1000 are content types or distinct words, and only 114:000 were function types. That is, there are many more content types than function types. Furthermore, there is much more repetition of function tokens than of content tokens.

A close look at the inventory of words present in the text reveals that 23% of the content words are repeated. That is, they appear more than once in the text either with the same conceptual structure (repetition of exactly the same form) or with a conceptual structure derived from the basic category. For example, the words *Petrobras* and *petróleo* (=crude oil) or *produtos*, *produzir*, *produção* (products, to produce, production), or yet *integrado* (to integrate, past participle, s.m.), *integridade* (integrity), *integrar* (to integrate) and *ambientais* (environmental; pl.), *ambiental* (*environmental*, s.), and *ambiente* (environment).

Most content tokens that are repeated more than once relate to the topic of discourse: Petrobras' efforts to integrate environmental protection, industrial safety, and the quality of its services and goods. Namely, *ambiente* (=environment) has 7 repetitions; *Petrobras*, the main actor and source of energy has 5 repetitions; *qualidade* (=quality), *produção* (=production), and *integrar* (=to integrate) have three repetitions; *segurança* (=safety), *industrial* (=industrial), *garantir* (=to guarantee), and *companhia* (=company) have

two repetitions. The other word that has two repetitions, *processo* (=process), is one of the few that enter in collocations. It collocates with Petrobras' efforts to integrate the three areas. Another collocation among the content words occurs between the words company and Petrobras. BP seems to use collocation of content tokens scarcely.

Likewise, BP generally does not rely on the situational context to implicate the actor or the figure of an argument. Rather, it repeats the entities. See the case of *Petrobras*, the main participant of discourse:

Petrobras reuniu a coordenação destas áreas...
Petrobras brought together the coordination of these areas

num único órgão... ligado diretamente à Presidência [da empresa]
under one single division answering directly to the chairman (of the company)

procurando [Petrobras] integrá-las à missão principal da Companhia...
trying [Petrobras] to integrate them to the mission main of the Company...

Este esforço integrado...denomina-se Processo Petrobras de Meio Ambiente....
This effort integrated is called Process Petrobras of Environment...

A Companhia busca se antecipar nas questões ambientais...
The Company tries to itself stay ahead in the issues environmental...

Para melhor conduzi-los, a Petrobras realiza diagnósticos ...ambientais
To better conduct them Petrobras undertakes diagnostic surveys environmental

In only three cases *Petrobras* is implicated: two in the situation context and one in the situational model. The other six mentions, as I said before, are explicit. Given that *Petrobras* appears as the actor in three of the five sentences the text has, it remains active during most of the time, either through explicit mentions or through morphological trappings in the verb.

The few instances in which it is not active, it is semi-active as a consequence of contextual effects (internal and external). That is the case of the second paragraph in which there are two sentences and *Petrobras* only appears in a head-modifier sequence as a modifier (e.g. *This integrated effort calls itself Petrobras Process of Environment, Quality and Safety*). However, it is implicated in the situational model.

Furthermore, the other entities that compose the topic of discourse remain active in the second paragraph through repetitions of the same conceptual structure (e.g. the words *integrado, integridade, meio ambiente, produção, produtos, qualidade* are repeated). In the third and concluding paragraph, only the environmental theme remains active through repetitions:

A Companhia busca se antecipar nas questões ambientais,
The (f.) Company endeavors itself to stay ahead in the (f.pl.) matters environmental

mantendo permanente diálogo com as comunidades afetadas
keeping on going dialogue with the (f.pl.) communities (f.pl.) affected (f.pl.)

por seus empreendimentos.
by its ventures (m.pl.)

Para melhor conduzi-los, a Petrobras realiza diagnósticos
To better conduct them (m.pl.) the (f.s.) Petrobras undertakes diagnostics

e monitoramentos ambientais que,
and monitoring (m.pl.) environmental (pl.) that

além de subsidiarem os Estudos de Impacto Ambiental (EIA),
besides subsidize (3rd. p.pl.) the (m.pl.) Studies (m.pl.) of Impact Environmental (SIE)

resultam em benefícios para a própria sociedade,
result (3rd.p.pl.) in benefits to the (f.s.) proper (f.s.) society (f.s.)

fornecendo novos dados sobre as condições ambientais
providing new (m.pl.) data (m.pl.) about the (f.pl.) conditions environmental (f.pl.)

da região e dos ecossistemas envolvidos.
of the (f.s.) region (f.s.) and of the (m.pl.) ecosystems (m.pl.) involved (m.pl.)

The environmental theme is repeated explicitly four times and collocate with *ecosystems* a fifth time. In BP, repetitions more than collocations seem to function as overt cohesive ties. Discourse entities remain active or semi-active most of the time requiring few cognitive operations toward the construal of a composite discourse lexicon schema.

By examining the grammatical classes in which the content words fit most frequently, nouns appear first in the rank (316:1000), immediately followed by verbs (158:000), adjectives (120:1000), and adverbs (13:1000). Apparently discourse forces drive morphology, creating texture by tying nouns, verbs, and adjectives together not only through the conceptual basis, but also through morphologic trappings. In other words, extensive development of discourse (observe the length of sentence number one in this text) occurs because morphological trappings allow nouns to connect to verbs and adjectives by means of trappings that are common to the three classes (person and number):

Consciente de que meio ambiente, segurança industrial e qualidade
 Certain (s.) that environment safety (f.) industrial and quality

são interdependentes, a Petrobras reuniu
 are interdependent (pl.) the (f.s.) Petrobras brought together (3rd.p.s; p.perf.)

a coordenação destas áreas num único
 the coordination of these (f.pl.) areas (f.pl.) (under one) single (m.s.)

orgão, a Susema, ligado diretamente à
 organization (m.s.) the (f.s.) Susema, connected (m.s.) directly to the (s.f.)

Presidência, procurando integrá-las à
 company chairman (s.f.) trying to integrate them (f.pl.) to the (f.s.)

missão principal da Companhia: produzir, comercializar
 mission (f.s.) main of the (f.s.) Company (f.s.): to produce, commercialize

e transportar petróleo e derivados para garantir
 and distribute crude oil and oil products (m.pl.) to guarantee

o abastecimento do País.
 the (m.s.) supplying (m.s.) of the (m.s.) country (m.s.)

Petrobras (s.) relates to the atemporal relation *consciente* (s.); *destas áreas* (f.pl.) relate to *meio ambiente, segurança industrial e qualidade* (the three areas; f.pl.), which, for their turn, relate to the modifier *interdependentes* (pl.), and so forth.

In many cases, discourse is extended by means of verbs that have lost the morphological trappings that are typical of their class (tense, mood, person) and appear as participles. The participles take trappings typical of modifiers (number and gender) since they enter phrasal structure as a metaphor of relative clauses. Consequently, participles carry back reference to the nouns that precede them. That is the case of *único* (m.s.) *orgão* (m.s.)...*ligado* (m.s.; past participle). An equivalent structure in English would be *a single division answering directly to the Chairman*. The English structure carries no cues, other than word order, that can tie the content words.

In other cases, discourse is extended by means of gerunds and infinitives. They generate non-finite subordinate clauses or right branching constructions that describe state-of-affairs. The non-finite clauses ground the independent clause, elaborating it and its head-noun. We will come back to this point when we discuss syntactic complexity in BP. This is the case of the verbs *procurando* (gerund), *produzir*, *comercializar*, *transportar*, and *garantir* (infinitives) in the example.

With regards to function tokens, 29% of them are repeated.. The number of contractions of the prepositions *de* (=of), *em* (=in), and *a* (=to) with the definite *o*, *os*, *a*, *as* (=the; masculine singular; m.pl.; feminine singular; f.pl.) and indefinite *um*, *uns*, *uma*, *umas* (=a or one; masculine singular; m.pl.; feminine singular; f.pl.) articles generate a sequence of repeated function types. You see forms like *nas* (*em+as*= in the; feminine plural), *num* (*em+um*= in the), *um*, *à* (to the; *feminine*). There are also contractions of the preposition *de* with the

deitics *estas* (=these [Neg.Awa]; closer to the addressee in space and time and far from the speaker; feminine and plural) and *esse* (=these [Spk.]; closer to the speaker; masculine and singular). Other forms that appear repeated times are *de*, *da*, *das*, *do*, *dos* etc.

The two most used tokens were the definite article *a* (=the; 82:1000 and its plural form) and the preposition *de* (=of; s.f.; s.m.; pl.f.; pl.m.; 76:1000 and its inflected forms), followed by the conjoiner *e* (=and; 51:1000). That articles rank first would be expected since among content tokens, nouns enjoyed most prominence and nouns are generally preceded by an article in BP.

In terms of grammatical classes, the most used function tokens were first, prepositions (95:1000), particularly *de* (=of) and its inflected forms; second, the definite article *a* (=the; 82:1000) with its inflected forms, and contractions (82:1000); third, pronouns (57:1000; among them 19:000 clitics; 19:1000 relative pronouns ;6:1000 deitics; 6:1000 reflexive personal pronouns; and 6:1000 possessive pronouns); and fourth 51:1000 conjoiners (*e*=and). Furthermore, it called my attention that no personal pronouns and a scarce number of possessive pronouns appeared. Personal and possessive pronouns, as the literature review demonstrated (see Chapter II), symbolize involvement, interaction with the audience. This would apparently indicate that BP has an informational rather than an interactional orientation. However, given that second mentions of personal pronouns are generally encoded in the verb, this interpretation needs to be reevaluated. Perhaps, BP texts simply have a different way of interacting with the audience.

Prepositions, in particular, appear in great number because they are part of verbs' sub-categorization or noun's sub-categorization:

<i>reuniu</i>	<i>a coordenação das áreas</i>	<i>num (em +um)</i>	<i>único órgão</i>
brought together		in one	

ligado *diretamente à (a, preposition + a, article) Presidência*
answering **directly** **to**

procurando integrá-las à (a, preposition + a, article) missão...
trying **to integrate** **them** **to the** **mission**

atender às (a, preposition + as, artigo) necessidades do presente
meet **to the** **needs** **of the** **present**

resultam em benefícios
result **in** **benefits....**

compatibilizar a produção com o meio ambiente
reconcile **the production of goods** **with** **the** **environment**

mantendo diálogo com
maintaining **dialogue** **with**

afetadas por seus empreendimentos
affected **by** **its** **ventures**

benefícios para a própria sociedade
benefits **to** **the** **proper** **society**

and because in BP autonomous entities typically appear first. Head-complement relations are frequent. For example, to express possessive relations BP draws on head-complement relations and allows no other type of relation (it cannot be marked morphologically like in English and it cannot be marked by the juxtaposition of postverbal nominals):

para garantir o abastecimento do (de+o) País
to (sign of an infinitive) **guarantee** **the** **supplying** **of the** **country**

assegurar a integridade do homem
safeguard **the** **integrity** **of the** **mankind**

necessidades do presente
needs **of the** **present**

This examples demonstrate how BP draws on head-complement relations, using prepositions where English would juxtapose typical modifiers or nominals that function as modifiers to elaborate a head in a syntagmatic relationship (BP syntax does not allow other combinations given that BP is a head first language). Other examples would be:

Estudos de Impacto Ambiental
 Studies of the Impact Environmental
 Environmental impact studies

dados sobre as condições ambientais da região e dos ecossistemas
 data about the conditions environmental of the region and of the ecosystems
 Environmental conditions data

Therefore, the number of prepositions in BP is directly linked to the structural characteristics of the language. Actually, this was one of the reasons that determined my decision to count only sequences of two or more prepositional phrases as marks of syntactic complexity. We will address this issue in the discussion of syntactic complexity.

In brief, this qualitative analysis demonstrates that to cue comprehension with the lexicon, BP relies mostly on repetition of content words and encoding of several grammatical functions in one token. It rarely relies on collocation of content types. Content types relate to each other through morphological trappings and clitics. The trappings with the clitics establish a net of landmarks that keep the trajectors under the focus of attention. Because tokens encode different grammatical functions, pronouns are scarce and NPs are rarely mentioned a second time. The most used function tokens were prepositions. That seems natural given that BP is a head first language and uses head-complement relations with high frequency. Finally, no personal pronouns, few possessive pronouns, and few locative and time adverbials were found in this qualitative analysis. BP texts seem to be primarily informational, an evaluation that needs to be re-addressed given the function of morphology in Brazilian Portuguese texts.

How does syntax cue meaning? The most striking syntactic fact about the sample text is the scarce number of sentences. The three paragraphs are formed by five sentences with 48, 36, 15, 18, and 41 words. The average number of words per sentence (=32) tells

us that sentence length is not an issue for the construction of meaning in BP. It also indicates that right branching constructions are quite common, and do not impact clarity.

Given the results of the quantitative analysis, one type of right branching construction we would expect to see frequently are prepositional phrases. However, this does not seem to be the case of the sample text. There are only two cases of juxtaposed PPs (15:000):

a coordenação destas áreas num único órgão

dados sobre as condições ambientais da região e dos ecossistemas
 data about the conditions environmental of the region and of the ecosystems

The PPs determine the profile of the autonomous entity *dados* (=data) and *coordenação* (=coordination), and establish layers of e-sites that are conceptualized sequentially. Observe the case of *dados*:

DADOS

sobre as condições ambientais

da região

e dos ecossistemas

The dependent entity about the environment conditions carries in its profile the e-site that fine grains the entity data. The two other PPs specify SITUATION [SPACE, NOTIONAL] of the environmental conditions under focus and establish an atemporal relation with the NP environmental conditions that, for its part, relates to data. The sequential scanning of the three

entities construe the final composite structure of the head-complement relation. As complements, PPs ground the figure that they follow, requiring no further conceptual rearrangement to construe the final composite structure.

Thus, adding to results of the quantitative analysis, sequences of PPs in this text do not seem to be the strategy BP writers use to extend discourse or sentence length. There seems to be other factors at play. A close look at the first sentence of this text (an action chain with 48 words) reveals that non-finite subordinate clauses appear to be the main means of topic development and, ultimately, to be responsible for lengthy sentences in the example:

Consciente de que meio ambiente, segurança industrial e qualidade são interdependentes,
 Certain that environment safety industrial and quality are interdependent

a Petrobras reuniu a coordenação destas áreas num único órgão, a Susema,
 Petrobras brought together the coordination of these areas into a single division, Susema,

ligado diretamente à Presidência,
 connected directly to the Chairman,

procurando
 trying

integrá-las à missão da Companhia:
 to integrate them to the mission of the Company:

produzir, comercializar e transportar petróleo...
 to produce, commercialize and transport crude oil

para garantir o abastecimento do País.
 To guarantee the supplying of the country

The sentence opens with a subordinate clause that is immediately followed by an independent clause in which the main actor, Petrobras, appears at clause initial position.

Then, a sequence of four non-finite nominal clauses follow. The first *ed*-participle clause indicates the result of the temporal process and accumulates the functions of a verb with that of an adjective. Similar to a verb, it expresses aspect and concludes the action of the main verb, brought together three areas in one division. Similar to an adjective, it is inflected for gender and number:

órgão (m.s.)... *ligado* (m.s.)
division that answers ...

Cognitively, the *ed*-participle clause is a landmark of division, the entity that is moving through discourse and that is being modified by that answers.

The second is a *ing*-participle clause whose verb (=procurando) expresses a temporal on-going process that took place after the main verb (=reuniu) and that modifies the main verb:

Petrobras reuniu as três áreas.
Petrobras brought together the three areas. (How?)

Procurando.
Trying

integrá-las à missão da Companhia
to integrate them to the mission of the Company.

Therefore, the *ing*-participle functions as an adverb and could be paraphrased by a coordinated clause that would begin with the coordinating conjunction and: Petrobras brought together the three areas and tried to integrate them to the mission of the Company. Cognitively, the *ing*-participle is a landmark.

Actually, to integrate them to the mission of the Company is a *to*-infinitive clause that complements the verb *procurando* (=trying), constituting its direct object. Though the conceptual structure of the infinitive carries the potential of a [PROCESS], it is in fact closer to a noun.

The fourth non-finite clause relates to the head-modifier sequence *a missão principal da Companhia* (=the Company's main mission):

produzir, comercializar e transportar petróleo e derivados
to produce, commercialize and distribute crude oil and oil products

Again, we have a *to*-infinitive clause that explains the mission of the Company, functions as a dependent entity, and is profile determinant.

Finally, the last landmark is also an infinitive clause: *para garantir o abastecimento do País* (=to safeguard the supplying of the country). It conveys a potential final consequence of the action expressed by the previous *to*-infinitive clause. This consequence could be paraphrased by *the Company's mission, producing, marketing, and distributing crude and oil products, will guarantee the supplying of the country*. Therefore, *para garantir o abastecimento do País* functions as an adverb, modifying the *to*-infinitives in the previous clauses in their roles of potential [PROCESSES].

In addition to lengthening its sentences by means of non-finite nominal clauses, BP discourse frequently intrudes these non-finite clauses as well as restrictive relative clauses between the actor of the independent clause and the temporal process that relates to it. In the sample text, this effort (the subject) is set apart from is called (the temporal process that relates to the subject) by a restrictive relative clause and by three non-finite clauses (see in bold):

Este esforço *integrado*,
This effort (m.s.) *integrated* (m.s.)

que pretende
 that *intends* (m.3rd.p.s.)

assegurar a integridade do homem,
to safeguard the integrity of mankind

compatibilizar a produção com o meio ambiente
 reconcile the production of goods with environment [protection]

e garantir a qualidade de produtos e serviços
 and guarantee the quality of products and services

denomina-se *Processo Petrobras de Meio Ambiente, Qualidade e Segurança Industrial.*
calls (3rd.p.s.) itself *Process Petrobras of Environment [Protection], Quality and Safety Industrial*

Despite the intrusion of the subordinate clauses, there is no ambiguity. Verb morphology and the clitic *se* (see underlined) keep the referent *esforço* active.

The tendency to elaborate on the topic through right branching constructions such as non-finite clauses may also be seen in the last paragraph:

(FIRST SENTENCE OF THE THIRD PARAGRAPH)

A Companhia busca se antecipar nas questões ambientais,
The Company tries itself to stay ahead of the matters environmental

mantendo permanente diálogo com as comunidades afetadas
keeping constant dialogue with the communities affected

(SECOND SENTENCE OF THE THIRD PARAGRAPH)

Para melhor conduzi-los,
In order better to conduct them

a Petrobras realiza diagnósticos e monitoramentos ambientais que,
the Petrobras undertakes diagnostic and surveys (pl.) environmental that,

além de subsidiarem os Estudos de Impacto Ambiental,
besides subsidizing (3rd.p.p.) the Studies of Impact Environmental

resultam em benefícios para a própria sociedade,
result (3rd.p.pl.) in benefits to the itself society,

fornecendo novos dados sobre as condições ambientais da região...
providing new data about the conditions environmental of the region

There seems to be no doubt that BP relies extensively on non-finite nominal clauses to lengthen its sentences and give continuity to the topic. Because the non-finite clauses function as landmarks that specify the autonomous action chain (the trajector), a linear interpretive ordering is possible, calling on no further mental rearrangement. In addition, the content types and the figures in the non-finite clauses generally connect to the trajector through verb and nominal morphology as demonstrated (see underlined in the example).

On the other hand, though the present example frequently uses non-finite clauses to develop the topic and relies limitedly on juxtaposed PPs, other pieces of discourse in the BP corpus illustrate well the quantitative results. That is, it is common for Brazilian Portuguese discourse to elaborate its nominal heads by means of sequences of PPs. In most texts in the corpus there is at least one long sequence of juxtaposed PPs. Take the next examples. In example (1), there are eight juxtaposed PPs (see in bold). In example (2) there is a sequence of 13 PPs:

(1) *Incentivo à Vida Natural* (=Lending Nature a Hand; Petrobras, 1994):

A Petrobras vem desenvolvendo projetos
 Petrobras has been developing projects

de incentivo
 of incentive

à vida natural
 to the life natural

à preservação
 to the preservation

de espécies ameaçadas
 of species endangered

ao esporte
 to the sports

ao ar livre
 outdoors

AND

à qualidade
 to the quality

de bens e serviços
 of goods and services

Alongside in-house programs in environmental preservation, Petrobras sponsors external projects to protect endangered species and fosters respect for our natural habitat. Other projects encourage participation in open-air sports and commitment to quality goods and services.

(2) *A Preocupação com o Meio Ambiente* (=Concern with the environment; Petrobras, 1994):

O IMA/AM é também responsável
The IMA/AM is also responsible

pela emissão
for the issuing

das licenças prévias
of the permits previous

para as atividades
to the activities

de pesquisa e exploração sísmica
of surveying and exploration seismics

[das] licenças
[of] permits

de instalação
of installation

para as atividades
to the activities

de perfuração
of drilling

de poços
of wells

E (AND)

[das] licenças
[of] permits

de operação
of operation

para cada campo produtor
to each well producer

de petróleo
of crude oil

IMA/AM is also responsible for issuing pre-authorizations for seismic exploration and surveying as well as for issuing installation permits for well-drilling activities and operating licenses for each oil-producing field.

These two texts, as well as many others in the BP corpus, developed its sentences mainly by means of sequences of PPs (there are seven long sequences similar to the ones exemplified above and eighteen with four juxtaposed PPs across the cases). Therefore, BP favors both PPs and non-finite clauses to develop its topics. Future quantitative research should code and tally non-finite clauses to check which of the two events is the most frequently used as a right branching construction.

With regard to attention units, in BP, the stretches of language between punctuation marks reflect, in general, the clause structure. If you examine the clauses illustrated below (each clause has been closed with double slash-- //), you will see that clauses frequently coincide with attention units. Punctuation seems to do part of the reader's job, parsing sentences into clauses:

- 1 *Consciente de que// meio ambiente, (5 words)*
- 2 *segurança industrial e qualidade são interdependentes, //(6)*
- 3 *A Petrobras reuniu a coordenação destas áreas num único órgão, (10)//*
- 4 *a Susema, //(2)*
- 5 *ligado diretamente à Presidência, //(4)*
- 6 ***Procurando // integrá-las à missão principal da Companhia: (8)//***
- 7 *produzir, (1)//*
- 8 ***comercializar // e transportar petróleo e derivados para garantir o abastecimento do País. (12)//***

- 9 *Este esforço integrado, (3)*
- 10 *que pretende assegurar a integridade do homem, (7) //*
- 11 ***compatibilizar a produção com o meio ambiente //e garantir a qualidade de produtos e serviços//, (15)***
- 12 *denomina-se Processo Petrobras de Meio Ambiente, (7)*
- 13 *Qualidade e Segurança Industrial. //(4)*
- 14 ***Um dos objetivos desse processo é// atender às necessidades do presente// sem comprometer o futuro. (15 words)***

- 15 ***A Companhia busca // se antecipar nas questões ambientais, (8)***
- 16 ***mantendo permanente diálogo com as comunidades // afetadas por seus empreendimentos. (10)***
- 17 *Para melhor conduzi-los, (4) //*
- 18 *a Petrobras realiza diagnósticos e monitoramentos ambientais que, (8) //*
- 19 *além de subsidiarem os Estudos de Impacto Ambiental (EIA), (9) //*
- 20 *resultam em benefícios para a própria sociedade, (7) //*
- 21 *fornecendo novos dados sobre as condições ambientais da região e dos ecossistemas envolvidos.(13)//*

GARANTIR
o abastecimento
do país

ASSEGURAR
a integridade
do homem

ESTUDOS
de Impacto
ambiental

GARANTIR
a qualidade
de produtos
e serviços

ATENDER
às necessidades
do presente

NOVOS DADOS
sobre as condições ambientais
da região
e dos ecossistemas
envolvidos

Because head-complement and head-modifier relations appear in the order they are conceptualized (first the autonomous and then the dependent entity), BP events optimize cognitive routines. Figure 18 illustrates the conceptualization of a head complement relation *estudo de impacto ambiental* (= study of impact environmental) within which there is a head-modifier sequence in BP. In English, the same syntagmatic relation would be a head-modifier sequence in which the profile determinant entity appears last—ENVIRONMENTAL IMPACT STUDY and, thus, requires conceptual re-arrangement.

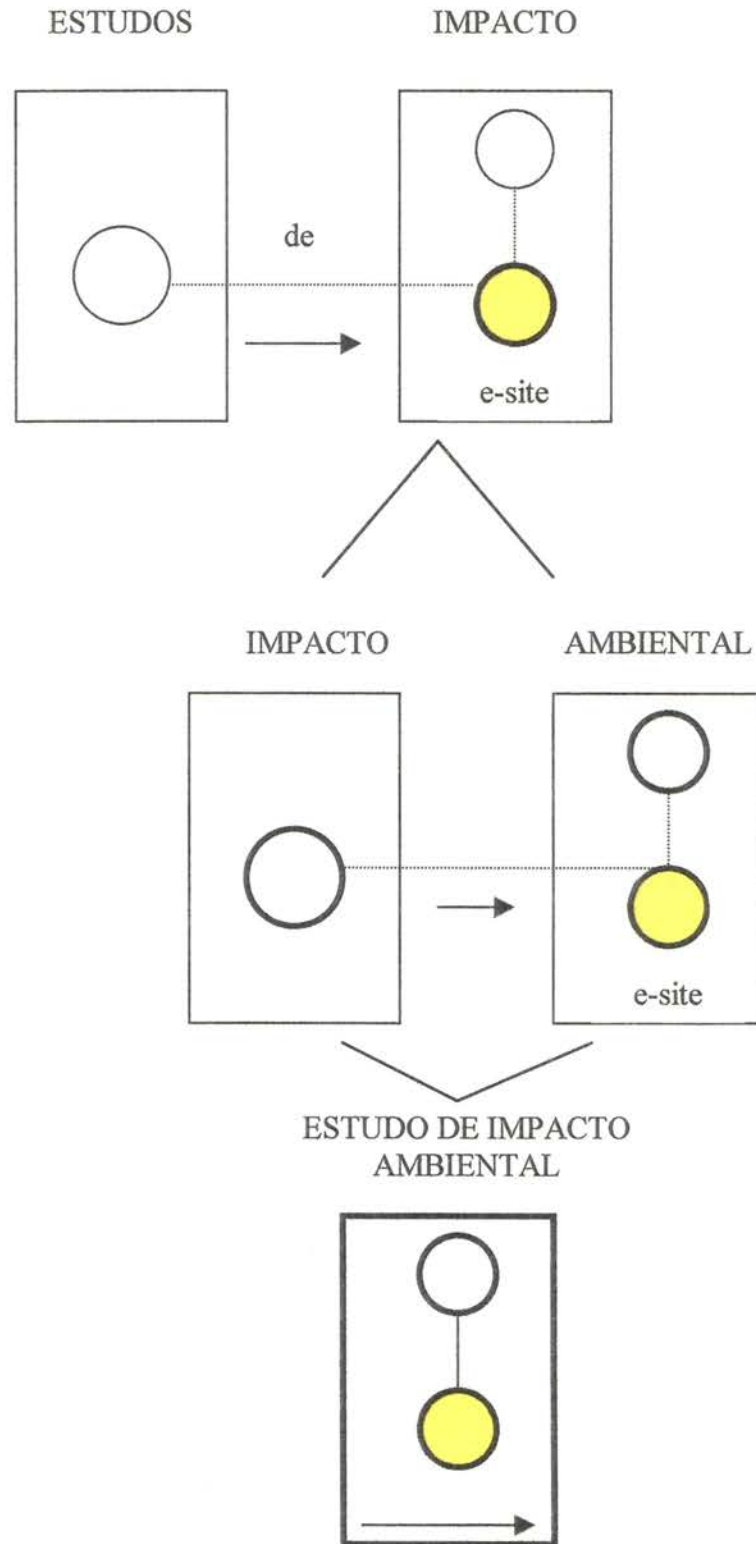


Figure 18. Conceptualizing syntagmatic relations in Brazilian Portuguese.

Contrary to the number of dependent clauses we find in this text, and similar to the quantitative results for the BP data, there are only 32:1000 independent clauses or autonomous action chains:

*A Petrobras reuniu a coordenação destas áreas num único órgão
Este esforço denomina-se Processo Petrobras de Meio Ambiente
Um dos objetivos desse processo é
A Companhia busca se antecipar nas questões ambientais
A Petrobras realiza diagnósticos e monitoramento ambientais*

All of them collocate with the topic. If we compare the number of independent clauses to the number of dependent clauses, we conclude that BP prefers to develop the topic by means of non-finite nominal clauses. In the cases in which independent constructions are used to cue meaning through the figure-ground asymmetry, the mental re-arrangement of the final composite structure is optimized through the morphological trappings.

The last attributes that we would expect to contribute to syntactic complexity are nominalizations. Contrary to expectations, in this piece of discourse there are only 38:1000 nominalizations: *coordenação* (=coordination), *presidência* (=presidency), *abastecimento* (=supplying), *produção* (=production), *empreendimentos* (=ventures), and *monitoramentos* (=monitoring). The three nominalizations that appear in the first paragraph--*coordenação* (=coordination), *presidência* (=presidency), *abastecimento* (=supplying)--yield maximal cognitive effects through text external sources. That is, the words *coordination* and *presidency* are part of business language and are used highly frequently. The word *abastecimento* is also very typical in the oil industry jargon. When

it is activated, users know it relates to the process of supplying consumers with gasoline. The processes these nominalizations reify enjoy prominence in readers' immediate reality by establishing correspondences with the oil industry business. Had the processes been spelled out, readers would have had to follow more cognitive routes than their knowledge of the situational model made necessary. In other words, it would have been a violation of the Principle of Relevance.

The same holds true for the nominalizations *produção* (=production) in the second paragraph and *empreendimentos* (=ventures) in the third paragraph. In addition, the nominalization *empreendimentos* functions as a point of reference, contributing to the development of the paragraph. Together with *monitoramentos* (=the checking of environmental conditions), they are further referred to in discourse by means of morphology, the clitic *-los* (in the case of *empreendimentos*), and the elided relative pronoun *que* (in the case of *monitoramentos*). The nominalization *monitoramentos*, however, does not yield the same contextual effects as the other nominalizations in the text. It is a rare word and reifies processes that are not so commonly known to every potential reader of the magazine in which the text appears. Thus, nominalizations in this text were few and most of them were maximally efficient. Reasons for such variation within the BP corpus as well as other variations we found in the corpora will be addressed under the *Variation within the COMMUNICATIVE TEXT* sub-heading (see next, following the qualitative analysis of BP examples).

To summarize, syntactic attributes such as long sentences organized in clause-like attention units, morphological trappings that keep figures in discourse active or at least semi-active most of the time, non-finite nominal clauses, juxtaposed PPs, head-modifier

relations and head complement relations that favor the order of conceptualization combine to depict the syntactic organization of the “communicative” BP institutional discourse.

How does BP institutional texts interact with reality? Based on the number of Adv.Ps. of time and place in this example, we would say that there is no interaction with reality. This lack of interaction has also been reflected by the lack of personal pronouns and the few possessive pronouns and deitics that appear in this piece of discourse. As I have mentioned before, it appears that BP institutional expository discourse is basically informational. If that is so, the situational model plays a fundamental role in activating contextual effects brought to bear by other content types in the text. This is the case of some nominalizations we have just analyzed.

However, though there are no Adv.Ps. of time and place to cue GOAL, ORIENTATION, and ALIGNMENT, BP uses position in the sentence and subordination to do the same job. The first paragraph opens with a subordinate appositive clause that explains why Petrobras took measures to bring together the environmental protection, industrial safety, and quality together under one division:

Consciente de que meio ambiente, segurança industrial e qualidade
 Certain (s.) that environment safety (f.) industrial and quality

são interdependentes, a Petrobras reuniu
 are interdependent (pl.) the (f.s.) Petrobras brought together (3rd.p.s; p.perf.)

a coordenação destas áreas num único
 the coordination of these (f.pl.) areas (f.pl.) (under one) single (m.s.)

orgão, a Susema, ...
 organization (m.s.) the (f.s.) Susema,

The writer placed that nominal clause in a perceptually salient sentence position. He could have moved it to the position right after *Petrobras* (*Petrobras*, certain that environmental protection, industrial safety, and quality are interdependent) or even to sentence-final position. The modifier *consciente* (=certain) connects it to *Petrobras* no matter what because of *Petrobras*' agent role. A paraphrase of the structure would be: Because *Petrobras* is certain that environmental protection, industrial safety, and quality are interdependent, it brought together the three areas.

However, the subordinate nominal clause comes first to be perceptually salient. In being perceptually prominent, it establishes the ORIENTATION and GOAL of discourse. Readers know, from the situational model, that *Petrobras* is the main actor and source of energy. It is inferrable given the publication in which the text has been published and its institutional GOAL. Therefore, perceptually, it is advantageous to the addresser to make its beliefs prominent, indicating the discourse ORIENTATION and GOAL up front.

When readers see the appositive clause, they already know the ORIENTATION is that of *Petrobras*, but they learn that discourse will talk about the interdependency of the three areas and about what types of actions *Petrobras* is taking to demonstrate the truth value of the belief the clause expresses.

In addition, ORIENTATION is further reinforced by subsequent mentions of the main actor—six explicit mentions and two mentions implicated by the situation context and the situational model.

With regard to ALIGNMENT, there is little explicit interaction with the audience. There are no moments in which the text explicitly establishes solidarity with the audience.

Certain that environmental protection....

This comprehensive effort to safeguard mankind...

The company endeavors to stay one step ahead in environmental matters...

Nevertheless, the text firmly defines the profile of Petrobras as an environment-friendly company through actions that in real life connect to a good environmental policy. Namely, to guarantee the quality of products and monitor gas emission; to integrate three areas that are critical to sustained development, etc. Action words in perceptually salient positions (clause-initial) corroborate that image:

CONSCIENTE....	(=Certain that...)
PROCURANDO...	(= Trying to...)
PRODUZIR.....	(= to produce)
COMERCIALIZAR	(=to distribute)
TRANSPORTAR	(=to transport)
COMPATIBILIZAR	(=to make compatible)
GARANTIR.....	(=to guarantee)
ASSEGURAR.....	(=to assure)
ATENDER	(=to meet the needs....)
FORNECENDO	(=providing)

The agent or source of energy is implicated by the situation context. This is a marketing strategy to negotiate role relationships with the audience. Therefore, the audience too is implicated in the situational model, in the situation context, and in the GOAL of discourse: to sell a positive image of Petrobras through its actions.

Summary. Qualitative analysis of one of the BP texts enlarged our understanding of how syntactic organization in BP optimizes comprehension. Among other things, the major devices BP used to develop the topic were clause-like attention units organized in lengthy sentences, morphology, juxtaposed prepositional phrases, and non-finite clauses. Furthermore, head-modifier and head-complement relations appeared in the order they are superimposed to construe the final composite structure, necessitating no further

mental rearrangement. The fact that attention units are clause-like and that word order favors the interpretive order counterbalanced the lengthy sentences in terms of comprehension. Nominalizations yielded cognitive gains in that they established correspondences with business language and the oil industry jargon or allowed further reference in discourse. Last, BP institutional expository texts seem to be basically informational if the criterion is the presence of AdvPs. Findings of this qualitative analysis indicate otherwise. Though there were no Adv.Ps. of place or time, subordinate clauses and action words placed at perceptually salient positions negotiated role-relationships with the audience. Besides, the situational model and the situation context also established the link with reality.

The results I presented revealed events associated with the English and the BP COMMUNICATIVE TEXT schemas and explained the cognitive routines readers follow to bind these events together in composite schematic structures. The co-occurrence of these events, as the qualitative analysis exposed, are motivated by cognitive effects, pragmatic and perceptual factors, and they constrain the representation of texts in each research condition. If the ORIENTATION-SOURCE-GOAL-DESTINATION schemas are controlled, when these events cluster, they form the complex gestalt or image-schema COMMUNICATIVE TEXT, an abstraction of the many cases that form the COMMUNICATIVE TEXT category in each condition.

Because the COMMUNICATIVE TEXT image-schema interacts with other schemas such as SCENARIO, MOTION, COMMUNICATION, PART-WHOLE, CONTAINER and with the conceptualizer of the texts, it gives rise to variations within the category or prototype effects. Next, I will briefly detail the range of instances of the

COMMUNICATIVE TEXT that appeared in the data and that clustered to form this abstract, simpler, and experience-grounded schema in the BP and English corpora respectively. Though these variations were not specifically addressed in the presentation of results, they were indicated in the tables that reported the distribution of the attributes across the data.

Variations within the COMMUNICATIVE TEXT schema

The variants or cases within the COMMUNICATIVE TEXT are conceptions of greater specificity and detail that elaborate the schema in contrasting ways. According to the results of the study, for example, *the English COMMUNICATIVE TEXT image-schema* is characterized by short sentences, long attention units, frequent independent constructions, collocation of content types, rare juxtaposed PPs or right branching constructions, and explicit interaction with readers' reality.

Within this abstract structure that encompasses commonalities of all members of the COMMUNICATIVE TEXT category, there is this more concrete, inclusive, and easier to recognize case which is the prototype. The prototypical English COMMUNICATIVE TEXT is characterized by a median number of 20 words per sentence, 10 words per attention units, 611:1000 content words, 53:1000 independent constructions, and 17:1000 juxtaposed PPs. This is the case this study characterized as typical and that is at the center of the category. It functions as a cognitive reference point because readers recognize it more readily and because it generates optimal processing.

Other examples of cases within the COMMUNICATIVE English TEXT category includes texts with an average of 34 words per sentence such as *Phillips' Company helps shape public policy*. If we examine this text, however, we see that geographic references

are the major responsible for sentence length. Namely, there are expressions such as *the Arctic National Wildlife Refuge* or *areas of the Outer Continental Shelf*. These references are in the situational model or SCENARIO schema, and yield positive cognitive effects. It also includes texts with 26 words per sentence such as Shell's *Existing Operations* or *Case Study: Improving Techniques in the Upstream* or texts with 16 words per sentence such as Phillips' *Safety*. Examination of Shell's *Existing Operations* reveal that if it were not for one sentence of 50 words, the average number of words per sentence in this text would be closer to the center of the category. This sentence consists of a list of issues that are part of exploration and production activities (the CONTAINER schema). Because most issues are sequences of nominalizations and a PP, the final sentence became lengthy:

The principal issues for the E&P industry are **disposal** of water produced with oil or gas; **utilisation** of associated gas produced with oil; **discharges** of contaminated drilling waste; and **management** of other industrial waste, all of which must be done in compliance with all applicable company standards and regulatory requirements.

Furthermore, there are variants within the COMMUNICATIVE English TEXT with BP-like attention units (=7 words long). For example, Phillips' *Company Takes New Management Approach*. These short units derive mainly from listings of corporate goals (= *safety, shareholder value, customer satisfaction, technology, employee satisfaction, technology, employee satisfaction and corporate citizenship*). Again, it is the interaction of the COMMUNICATIVE English TEXT schema with the PART-WHOLE schema that

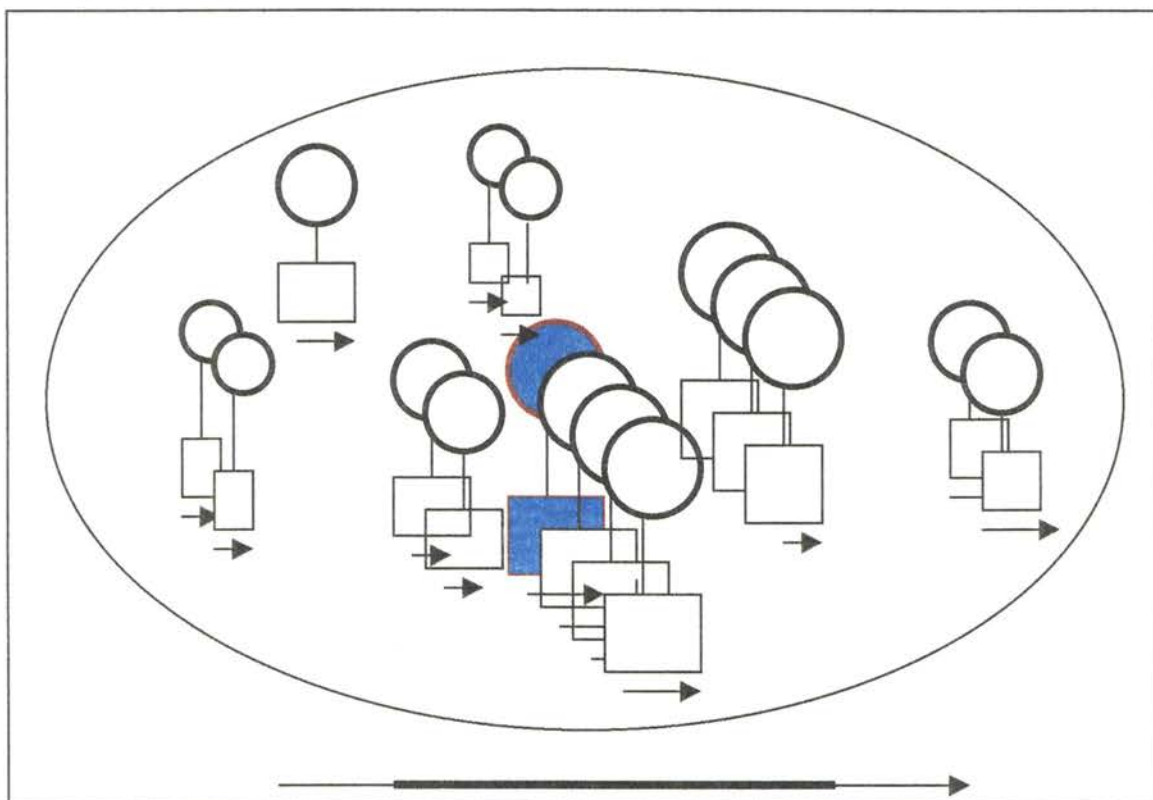
gave rise to this variant. Figure 19 diagrams the conceptualization of the English COMMUNICATIVE TEXT image-schema and its variations.

The diagram uses CG descriptive tools. Conceptualizers superimpose the variants or different cases and reify them into this abstract schematic structure of English texts. Because the SOURCE-GOAL-DESTINATION-ORIENTATION schemas are controlled, they are not represented in the figure. The only sub-component of the final structure represented is the PATH schema (see Chapter IV, Figure 5, for the complete diagram of the COMMUNICATIVE TEXT image-schema). The case highlighted in blue represents the central case.

The COMMUNICATIVE BP TEXT image-schema exhibits the same effects. The image-schema constitutes texts with long sentences; short attention units; frequent embedding and right branching constructions, specially juxtaposed PPs and non-finite clauses; absence of explicit actors; perceptual salience of autonomous entities; frequent nominalizations; repetition of content words; presence of morphologic trappings that keep autonomous entities active or semi-active; and implicit interaction with readers' realities.

Within this schema, there is a central case of the COMMUNICATIVE BP TEXT schema that is easier to process and recognize. Prototypically, BP texts are characterized by few-24-word long sentences; many seven-word long attention units; 65 nominalizations, 35 juxtaposed PPs, 40 independent constructions, and 553 content words in each thousand words; and a type-token ration of 0.46. BP discourse is also typically characterized by head-complement and head-modifier relations that favor the conceptualization ordering.

CASES



THE ENGLISH COMMUNICATIVE TEXT IMAGE-SCHEMA

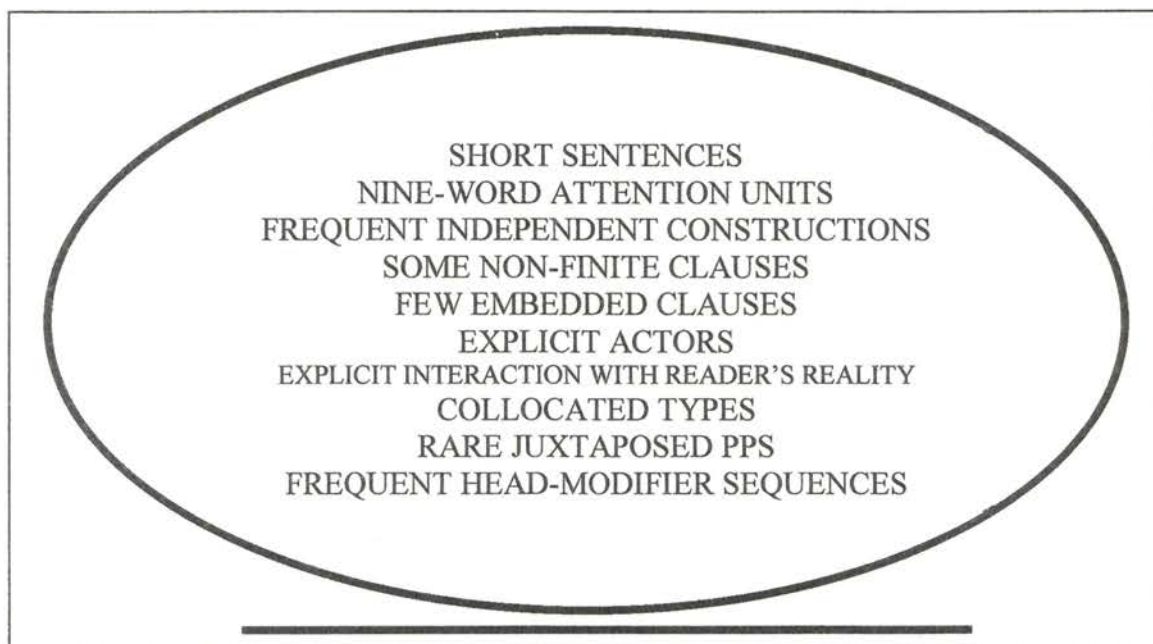
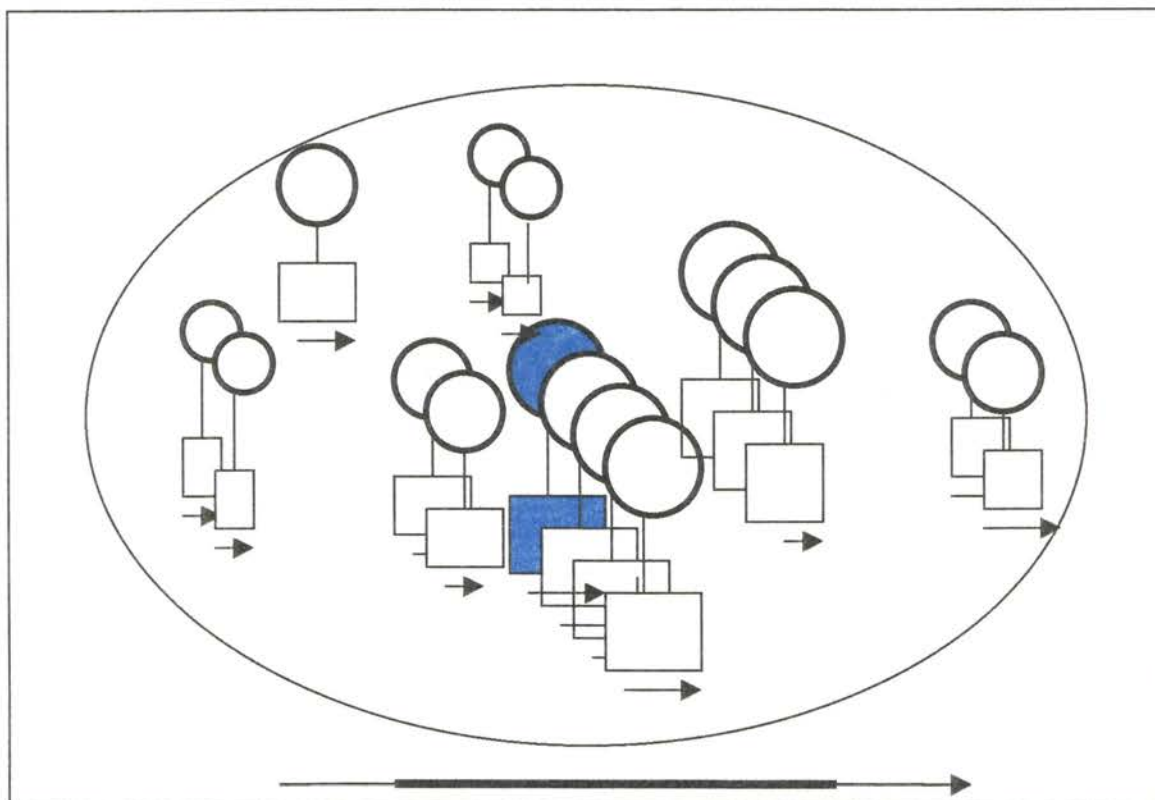


Figure 19. The English COMMUNICATIVE TEXT image-schema.

Other cases of the COMMUNICATIVE BP TEXT include texts with an extremely high number of nominalizations such as Petrobras' *A Perfuração e seus Desafios* (n=113:1000; Drilling and its Challenges). Examination of *A Perfuração e seus Desafios* reveals a high number of repetitions of the word *perfuração* (6 mentions; 14:1000). This word accounts for 56% of the nominalizations in the text and is the topic of discourse. If we consider that BP relies extensively on repetition to cue meaning, it is natural that *perfuração* was frequently repeated and that the occurrence of nominalizations increased. Another word that occurs three times in this text is *operação* (3 mentions; 7:1000). It is also related to the topic of discourse—the challenges of drilling operations. In brief, the high number of nominalizations in this case derives from an interaction with the COMMUNICATION schema.

On the other hand, there will be cases with a much lower number of nominalizations (n=38:1000) such as *Esforço para Preservar o Futuro* (=Striving to Safeguard our Future), the text treated by the qualitative analysis. Cognitively, the few nominalizations compensate for lengthier sentences (an average of 32 words per sentence). There is a real processing motivation for this variation. The interaction, in this case, is also with the COMMUNICATION schema. The conceptualizer wants the message to get to the destination with the least effort possible. Figure 20 diagrams these different cases and how they are superimposed to construe the BP COMMUNICATIVE TEXT image-schema. Similar to the diagram of the English COMMUNICATIVE TEXT image schema, Figure 20 reflects only the PATH sub-component of the final composite structure given that the other sub-components were kept constant. The case

CASES



THE BP COMMUNICATIVE TEXT IMAGE-SCHEMA

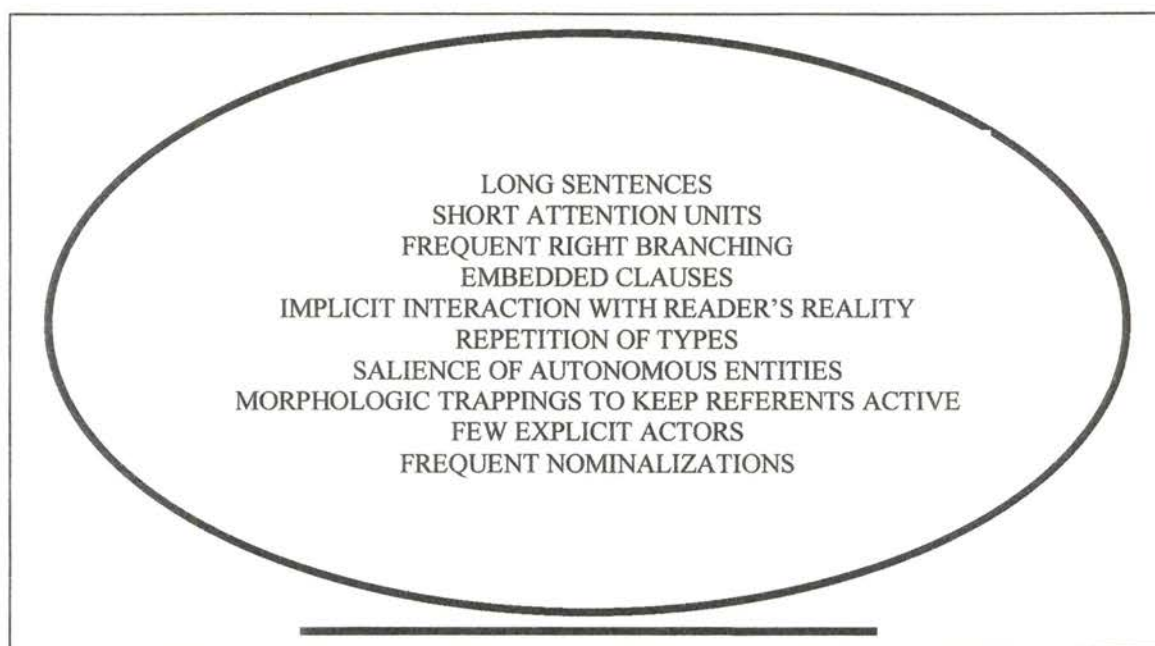


Figure 20. The BP COMMUNICATIVE TEXT image-schema.

highlighted in blue represent the prototype or the more inclusive member. In Chapter IV, the reader finds a complete diagram of *Texts as image-schemas* (Figure 5).

Also, in case there is need for further analysis of these variations, I direct the reader to the tables that display the distribution of the attributes across texts in each research condition. According to this study and its cognitive approach, the different cases (extremes and more central cases) within the BP and English text categories are conceptualized together on the bases of what they have in common. This understanding may be better explained by means of an example Lakoff (1990) uses. According to him, though the movements people use to iron clothes are never the same, they are all movements of a kind and, therefore, are categorized together despite their variations.

To examine if and which of the events investigated distinguish the English institutional discourse from the BP one, we need to compare medians and the counts of events per thousand words. We also need to check the statistical significance of any possible difference in the rank sums. Then, we should compare the results to those of the qualitative analysis. The next section will address such a comparison.

Comparing Brazilian Portuguese to English institutional discourse

This section compares the medians and counts of events as well as tendencies revealed in the previous section. In addition, it investigates if possible differences between the rank sums in each research condition are statistically significant by means of a two-tailed Mann Whitney-U test for independent groups. Comparisons of the qualitative analysis follow thereafter.

Table 26 lists the counts per thousand words, ratio or medians of attributes that I examined for the English and Portuguese research conditions. It also groups the

linguistic events according to the major linguistic categories they represent. That is, number of sentences, words per sentence, attention units, words per attention units, sequences of juxtaposed PPs, nominalizations, and independent clauses appear under *syntactic organization*. The type-token ration and the number of content words in each thousand appear *under lexicon organization*; and the number of locative/time adverbials under *interaction with concrete reality*.

TABLE 26
COMPARING BP TO ENGLISH INSTITUTIONAL DISCOURSE
COUNTS EXPRESSED PER 1,000 WORDS

	ENGLISH	BP
<i>Syntactic organization</i>		
Sentences	47	38
Number of words per sentence	20	24
Attention units	95	122
Words per attention units	9	7
Attention units per sentence	2	3
Sequences of 2 or + PPs	17	35
Sequences of 2 PPs	13	22
Sequences of 3 PPs	3	9
Sequences of 4 or + PPs	1	4
Nominalizations	61	65
Independent clauses/sentences	53	40
<i>Lexicon organization</i>		
Type-token ratio	0.50	0.46
Content words	611	553
<i>Interaction with concrete reality</i>		
Locative/time adverbials	37	37

Syntactic organization

Figure 21 illustrates the differences in the median number of words Brazilian Portuguese and English typically use in their sentences and attention units. Writers in Brazilian Portuguese combine more words into sentences than English writers. While BP sentences have a median length of 24 words, English sentences have a median length of 20 words. Furthermore, in English many more sentences group at the lower end of

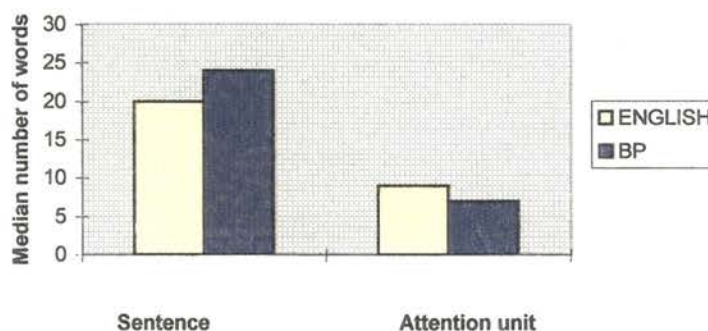


Figure 21. Sentence and attention unit lengths in BP and English discourse.

the range (15-24 words) while in Portuguese they group at the center of the distribution (17-31 words) with a slight tendency toward the upper end of the scale (32-55 words).

A Mann-Whitney U test on the rank sums of each sentence length in the two groups reveals that the difference between BP and English median sentence length is statistically significant (Mann-U test=26951; $p < 0.05$). Sentence length is an event that typically distinguishes the COMMUNICATIVE TEXT image-schema in English and BP institutional discourse.

The number of sentences per thousand words provides extra evidence for the ecological validity of sentence length in BP and English. BP uses lengthier but fewer sentences (38:1000) than English (47:1000).

If we examine the way in which these sentences are formed, we see that BP writers use more attention units ($n=122$) per sentence than English ($n=95$) writers. However, compared to English, those units in BP are shorter—a median length of seven words per attention unit. Furthermore, most attention units in the BP data stay within the bounds of 7 words. In English, on the other hand, the median length of an attention unit is nine words and 52.5% of them conform to the nine-word constraint. Other units (31%) vary from 10 to 17 words. The maximum number of words per attention unit in each research condition confirms the respective trends. While English has an upper limit of 42 words per attention unit, BP has an upper limit of 32 words.

According to results of the Mann Whitney-U test, the difference between the rank sums of attention units length in BP and English is also statistically significant (Mann-U test = 280425.500; $p < 0.05$). That is, we may be 95% confident that those results are not due to chance. Length of attention units is also an event that distinguishes the COMMUNICATIVE TEXT in English from the COMMUNICATIVE TEXT in BP.

Surprisingly, BP shows a moderate number (35:1000) of juxtaposed prepositional phrases, a device that typically increases the size of attention units. However, compared to English (17:1000), BP shows a significantly higher number of two or more juxtaposed PPs. In addition, while in English sequences of three juxtaposed PPs are rare (3:1000) and sequences of four (1:1000) rather unusual, in BP they appear more frequently. Actually, BP has sequences with five, six, even more juxtaposed PPs as exemplified in the qualitative analysis. In brief, the “communicative” BP institutional text tends to show significantly more juxtaposed PPs than its English counterpart. The ranges of each research condition themselves are quite revealing. While texts in BP have from 14:1000

to 54:1000 juxtaposed PPs, texts in English have from 0 to 29:1000. This tendency is statistically significant as confirmed by a Mann-Whitney U-test on the rank sums of each research condition (Mann-U test=14.500; $p<0.05$). BP and English also differ significantly in their use of two juxtaposed PPs (Mann-U test=62.000; $p<0.05$), three juxtaposed PPs (Mann-U test=53.000; $p<0.05$), and four juxtaposed PPs (Mann-U test=79.000; $p<0.05$). BP uses prepositional phrases to expand its attention units and develop the topic with a significant higher frequency than English.

Confirming my expectations, writers of the BP sample used nominalizations (65:1000) slightly more frequently than writers of the English sample (61:1000). Nevertheless, the difference is not statistically significant (Mann-U=177.000; $p>0.05$). The frequencies in each research condition have similar trends. Many cases concentrate at the center of the distribution. While in BP 45% of the texts display from 45-62:1000 nominalizations, in English 45% of the texts show from 40-67:1000 nominalizations. In addition, when the frequencies do not appear at the center, they tend toward the upper end of the scale. English and BP use nominalizations in similar ways.

With regard to number of independent sentences/clauses, while BP shows 40 independent clauses per thousand, English shows 53. The distribution of independent clauses in each condition confirms the different ways in which English and BP use this attribute. That is, whereas English institutional discourse exhibits 43 to 61 independent clauses per thousand words in 60% of the cases, pending toward the upper range of the scale, BP exhibits 31-39 independent clauses in 55% of the cases, pending toward the lower end of the scale. As I have demonstrated in the analysis of each research condition, the most frequent numbers of independent clauses and the maximum and minimum

values in each research condition provide extra evidence toward the different trends. Furthermore, the Mann Whitney-U test attests the statistically significant difference between BP and English discourses with regard to number of independent sentences or clauses (Mann-U test = 338.000; $P < 0.05$).

Figure 22 visually contrasts the number of sentences (bar chart # 1); number of attention units (bar chart # 3); total number of two or more PPs in sequence (bar chart # 5); sequences of two PPs (bar chart # 7); sequences of three PPs (bar chart # 9); sequences of four or more PPs (bar chart # 11); nominalizations (bar chart # 13); and independent constructions (bar chart # 15) per thousand of words in Brazilian Portuguese and English.

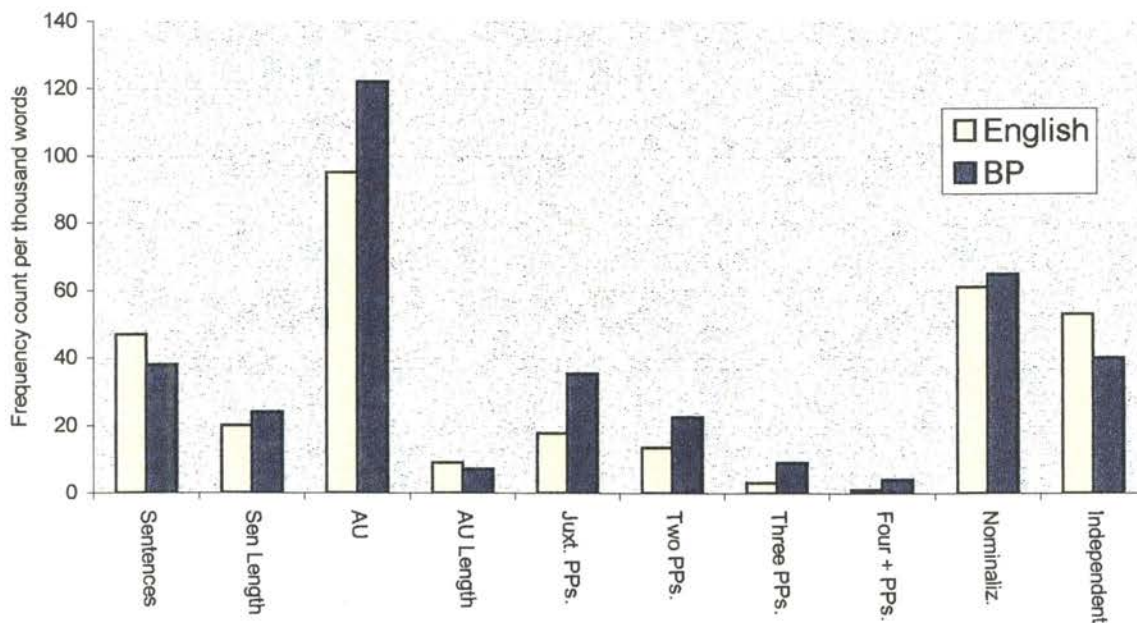


Figure 22. Comparing syntactic organization in English and BP.

Lexical organization

Figure 23 displays how BP and English go about using their lexical inventory. Bar chart number one represents the number of types and bar chart number two the number of content words in each thousand. Brazilian Portuguese repeats words with a slightly greater frequency than English. That is, English has a higher type-token ratio (=0.55) than Portuguese (=0.46). Also, English writers in the sample not only use a somewhat more varied vocabulary but choose words with greater precision than BP writers. In each thousand English words we find 611 that carry content as opposed to grammatical information whereas in each thousand BP words only 553 words that carry content information.

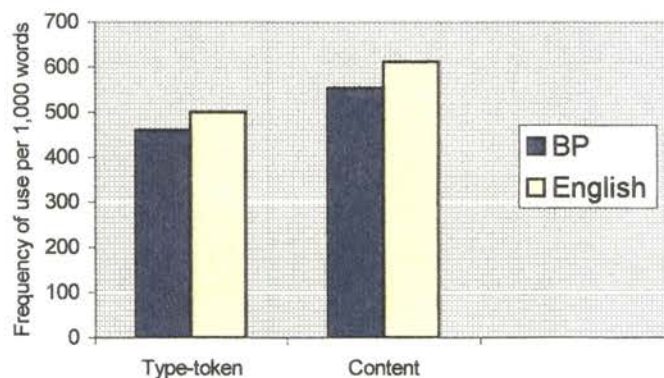


Figure 23. Lexical organization in English and BP institutional discourse.

With regard to the type-token ratio, 60% of the BP texts tend toward the lower end of the range (0.36 to 0.51) while 60% of the English texts tend toward the upper end of the range (0.52 to 0.74). With regard to the number of content words, though the tendency in BP institutional discourse is toward the upper end of the range

(550-590:1000; 75%), the range in English is much higher (the minimum in English corresponds to the median in BP = 554:1000). Furthermore, while 70% of the English texts (n=14) display more than 605:1000 content words, in BP only 30% of the texts exhibit this number.

To check if these results were due to chance, I ran the Mann Whitney U-test and found that though the difference for the type-token variable is not statistically significant (Mann-U = 244.000; $p > 0.05$), the difference for number of content words is significant (Mann-U= 392.000; $p < 0.05$). Precise lexical choice distinguishes the “communicative” English institutional discourse from the BP institutional discourse.

Interaction with concrete reality

English and Portuguese seem to draw equally on devices that would explicitly connect to the concrete reality of the readers. Both languages show 37 locative and time adverbials in each 1000 words. As expected, this dimension does not distinguish the two text categories (Mann-U=226.000; $p > 0.05$) statistically.

Table 27 reviews the results of the Mann Whitney-U test that examined the statistical difference between the rank sums of each attribute in the two research conditions as well as indicates the likelihood that the differences were due to chance. According to the quantitative results, sentence and attention unit length are the events that most typically distinguish the two research conditions.

Qualitative analysis of these attributes in the context of discourse fine grained our interpretation of the contrasts indicated by the quantitative analysis. For example, though the statistic test indicated that the numeric difference between the type-token ratio of the two language corpora is not significant, the way in which they use tokens in terms of

TABLE 27

COMPARING THE RANK SUMS OF ATTRIBUTES IN BP AND ENGLISH DISCOURSES

*p<0.05; df=1

	RANK SUMS		MANN-U test	P
	English	BP		
Sentence Length*	72402	74751	26951	0.000
Attention Unit Length*	468003.5	500132.5	280425.5	0.000
Sequences with 2 or + PPs*	224.5	595.5	14.5	0.000
Sequences with 2 PPs*	272	548	62	0.000
Sequences with 3 PPs*	263	557	53	0.000
Sequences with 4 or + PPs*	289	531	79	0.000
Independent clauses/sentences*	548	272	338	0.000
Content words*	602	218	392	0.000
Nominalizations	387	433	177	0.534
Type-token ratio	454	366	244	0.233
Locative/time adverbials	436	384	226	0.482

open and closed classes of words seem to differ. The analysis of one of the BP texts indicated that BP discourse draws on grammar with greater frequency than English. While the number of function tokens was the same for both languages, BP presented a greater inventory of function types than English. Furthermore, the English text repeated content types explicitly, implicated them in other instances, and collocated them with great frequency to develop the topic of discourse. The BP text, on the other hand, used morphology to keep referents active and develop discourse. Both languages repeated function words with greater frequency than content words, among them the, and, and of.

In addition, the texts in the two languages differed in terms of frequency of use of grammatical classes. While the BP text used prepositions and their inflected forms with the highest frequency, English used the definite article the and the coordinating conjunction and. Both languages used more content types than function types, though Portuguese used a slightly higher number of function types and a moderate lower number

Another difference in the way the two languages went about syntactic organization in the two texts the qualitative analysis addressed, relates to word order and conceptualization. The analysis of the BP text revealed that though the language draws on right branching constructions such as non-finite clauses and head-modifier relations, word order guarantees minimal mental re-arrangement of the entities involved in the correspondences. The autonomous entities appear first, allowing interpretation to follow the conceptualization order. English on the other hand, requires more mental re-arrangement given that in head-modifier relations the autonomous entity comes last. Therefore, readers conceptualize each entity and then need to re-arrange them to form the final composite structure.

The last attributes I qualitatively analyzed for the syntactic organization dimension were nominalizations. Findings of the qualitative analysis indicated that both research conditions use nominalizations in similar ways. They serve as reference points to further development of the topic and yield positive cognitive effects. However, while the BP text referred back to the nominalizations by means of function tokens (morphologic trappings, relative pronouns, and clitics), English did it by means of collocations. This finding seems to corroborate previous findings about the use of function and content tokens in the two languages.

Finally, though quantitative results indicated that BP and English interact with reality in similar ways, the qualitative analysis indicated slight differences. For example, the BP text seemed more detached from the context of discourse than the English text. It did not use personal pronouns or locative and time AdvPs at all. Furthermore, there were no moments in which the BP text explicitly established solidarity with the audience while

in English the very first paragraph establishes that solidarity. However, the BP text implicated that solidarity by means of the situational model, the situation context, and the GOAL of discourse. Thus, while English explicitly expressed ALIGNMENT, Brazilian Portuguese construed implicatures that allowed its inference.

Summary. A Mann Whitney-U test checked if the differences in the rank sums of the selected attributes differed significantly ($p < 0.05$) in the two research conditions. Results indicated that we may be 95% confident that the difference in the way BP and English use sentence length, attention unit length, juxtaposed PPs, content words, and independent clauses did not occur by chance. Those events are typical attributes of the COMMUNICATIVE TEXT in each condition. Qualitative analysis corroborate those findings indicating that while English uses content types to cue meaning, BP uses repetitions and grammar. Furthermore, while most attention units in BP seem to be clause-like units that right-branch by means of non-finite clauses, English attention units seem to be multi-clausal combinations of independent and dependent constructions. Other findings of the qualitative analysis that seem to distinguish the two research conditions include the order of conceptualization of head-modifier relations in the two languages. The head-first pattern of BP syntagmatic relations facilitates the construal of the final composite structure while English head-final pattern demands further rearrangement of the layers before the final structure is construed.

Explaining English translation

The results of the quantitative analysis in Table 28 portray English translation as a fuzzy category at the boundaries of English and BP institutional discourse categories. That is, English translation appears as a transformation or metaphor of the BP and

English image-schemas in terms of frequency per 1,000 words of the linguistic attributes I analyzed. For example, the median sentence length for English translation was 21 words (max.=71; min.=7) while BP texts yielded a median of 24 words and English texts a median of 20 words. If we look at number of sentences in each thousand words, we see that BP used 38 sentences to organize the image-schema while English translation used 45 and English 47. Translated discourse appears as a transformation of BP's schematic model toward the English schematic model. In that the conceptual structures of the source language are projected onto the English image-schematic model, we find more and shorter sentences in translated discourse than in BP, but not as many shorter sentences as we find in typical English discourse.

The same holds true for number and length of attention units. Translation discourse appears at the boundaries of the source and the target language COMMUNICATIVE TEXT categories. It uses fewer attention units per thousand words than BP (n=108 versus n=122 for BP texts), but more attention units than typical English discourse (n=95). Lengthwise those units are equivalent to BP units. They have a median of seven words (n=7) compared to nine words in English (n=9).

English translation reveals the same effects for other linguistic attributes I analyzed. It organizes and uses content words, independent constructions, sequences of four PPs in ways that resemble typical English discourse, but that are still similar to Portuguese.

However, with regard to total number of juxtaposed PPs per thousand words, number of two juxtaposed PPs, and number of nominalizations in each thousand words, English translation appears as a member of the COMMUNICATIVE English TEXT image schema. It uses these attributes with a slight lower frequency than typical English

discourse, but it is consistent with the trend established in the image-schema. Or, in the case of the type-token ratio, it uses the attribute with a slight higher frequency, drawing on an even higher range of vocabulary.

TABLE 28
ENGLISH TRANSLATION AS A METAPHOR OF ENGLISH AND BP DISCOURSES
PER 1,000 WORDS

Attributes	English	English Translation	BP
<i>Syntactic organization</i>			
Sentences	47	45	38
Median number of words per sentence*	20	21	24
Attention units	95	108	122
Median # words per attention units*	9	7	7
Attention units per sentence	2	2.4	3
Sequences of 2 or + PPs	17	15	35
Sequences of 2 PPs	13	10	22
Sequences of 3 PPs	3	3	9
Sequences of 4 or + PPs	1	2	4
Nominalizations	61	51	65
Independent clauses/sentences	53	51	40
<i>Lexical organization</i>			
Type-token ratio*	0.50	0.52	0.46
Content words	611	594	553
<i>Interaction with concrete reality</i>			
Locative/time adverbials	38	41	37

* When not indicated, the counts are expressed per thousand words. Overall counts for the corpora.

However, in some of those instances English translation does not differ statistically from the other two research conditions as Table 29 demonstrates. A Friedman two-way analysis of variance compared the mean rank sums of each linguistic attribute for the three groups and indicated that though English, English translation, and Brazilian

Portuguese differ numerically in the way they organize and use nominalizations (Friedman test (2)=3.900; $p>0.05$) and adverbs (Friedman test (2)=2.500; $p>0.05$), those differences carry no statistic significance. That is, those attributes are equally distributed in the ranking as far as the three corpora of the study or similar ones are concerned. English translation organizes and uses nominalizations, and time and locative adverbials in ways that resemble English and BP.

On the other hand, the same test indicated that writers in the three research conditions used attention units (Friedman test (2)= 27.158; $p<0.05$); sentence lengths (Friedman test (2)=19.550; $p<0.05$); type-tokens (Friedman test (2)=8.400; $p<0.05$); content words (Friedman test (2) =25.525; $p<0.05$); independent sentences or clauses (Friedman test (2) = 25.525; $p<0.05$); juxtaposed PPs (Friedman test (2) =27.175; $p<0.05$); sequences of two (Friedman test (2) =14.800; $p<0.05$), three (Friedman test (2) = 13.975; $p>0.05$), and four PPs (Friedman test (2) =7.525; $p<0.05$) in ways that differ statistically. That is, the differences among the ranks did not occur by chance ($p<0.05$).

A post-hoc Nemenyi test located precisely where the differences occurred. It revealed, for instance, that length of attention units distinguishes English translation from BP and English statistically. The difference between the mean rank sums in the matrix are higher than 0.13 (E-BP > Nemenyi=0.13; E-T>Nemenyi=0.13; and BP-T>Nemenyi=0.13). The average correlation among conditions is extremely low (Kendall coefficient of oncordance=0.022). While the distance of the English translation from BP does not surprise us at all, that is not true of the distance English translation takes from English discourse itself. Indeed we would expect the two types of discourse

TABLE 29
 COMPARING THE THREE RESEARCH CONDITIONS
 DF=2; *p<0.05

ATTRIBUTES	ENGLISH	ENGLISH TRANSLATION	BRAZILIAN PORTUGUESE	FRIEDMAN TEST	NEMENYI POST-HOC TEST	KENDALL
	Mean Rank Sum	Mean Rank Sum	Mean Rank Sum			
<i>Syntactic organization</i>						
Sentence length*	1.5	1.7	2.2	19.55	0.2	0.04
Attention unit length*	2.1	1.9	1.4	27.16	0.1	0.02
Sequences of 2 or + PPs*	1.6	1.5	3.0	27.17	0.77	0.70
Sequences of 2 PPs*	1.7	1.6	2.7	14.80	0.77	0.40
Sequences of 3 PPs*	1.8	1.6	2.7	13.97	0.77	0.35
Sequences of 4 or + PPs*	1.6	1.9	2.5	7.53	0.77	0.19
Nominalizations	1.9	1.8	2.4	3.90	0.77	0.09
Independent sentences*	2.3	2.5	1.1	23.73	0.77	0.60
<i>Lexical organization</i>						
Type-token ratio*	2.1	2.4	1.5	8.40	0.77	0.01
Content words*	2.6	2.2	1.1	25.53	0.77	0.64
<i>Interaction with reality</i>						
Locative/time adverbials	2.0	2.2	1.7	2.50	n.s.	--

NOTE: n.s.=non-significant or p>0.05; Nemenyi values were rounded to the second decimal place.

to use and organize attributes in similar ways. Again, English translation appears as a fuzzy category at the boundaries of BP and English image-schemas.

As expected, the Nemenyi test also revealed that length of sentences distinguishes only English and Brazilian Portuguese, and Brazilian Portuguese and English translation research conditions. For those pairs in the matrix the difference between the mean rank sums were higher than 0.2 ($E-BP=0.70 > Nemenyi=0.2$; $BP-T=0.5 > Nemenyi=0.2$). As for English translation and English, the Nemenyi attests no significant difference in the mean rank sums assigned to them ($E-T=0.2$). Actually, they are equivalent.

In the case of the type-token ratio, English translation yielded mean rank sums that differed significantly from Brazilian Portuguese ($BP-T=0.9 > Nemenyi=0.77$), but, as we would hope for, did not take distance statistically from English ($E-T=0.3 < Nemenyi=0.77$). Though the two conditions use types and tokens in similar ways, examination of the mean rank sums suggests that English translation is an extension of the English text category in that its type-token ratio is higher. The Kendall coefficient of concordance was low as well ($=0.210$).

The post-hoc test yielded similar results for content words. While English translation and BP differed significantly ($BP-T=1.1 > Nemenyi=0.77$), English and English translation appeared as members of the same family, yielding mean rank sums that do not differ statistically ($E-T=0.4 < Nemenyi=0.77$). The Kendall coefficient of concordance was 0.638.

Counter expectations, English translation used four or more juxtaposed prepositional phrases in ways that interacted with BP discourse

(BP-T=0.57<Nemenyi=0.77). However, English translation ways with four or more prepositional phrases also interacted with English discourse (E-T=0.27<N=0.77). This means that English translation is as close to typical English discourse as it is to typical BP discourse, and confirms the interpretation of English translation as a fuzzy category. According to expectations, though, English translation resembles English discourse in terms of total number of juxtaposed PPs and number of three and two juxtaposed PPs while it takes distance from Brazilian Portuguese discourse.

Finally, with regard to independent constructions, the mean rank sums assigned to English translation and BP (BP-T=1.4>Nemenyi=0.77) differ statistically while those assigned to English translation and English (E-T=9.2<Nemenyi=0.77) do not. English translation and English use independent sentences as members of the same COMMUNICATIVE TEXT category. The average correlation among the conditions as expressed by the Kendall Coefficient of Concordance was 0.593. In addition, examination of the mean rank sums suggests that English translation extend from the English text category in that it uses independent clauses/sentences with greater frequency. In conditions similar to those of this dissertation, we expect to find similar results.

Figure 24 provides a visual representation of how the discourse translation corpus interacts statistically with its source texts and to the English corpus. The bars represent the numeric difference among the mean rank sums for each attribute in English, English translation, and BP. The chart is organized in four parts: From left to right, it first illustrates attributes that English translation uses in ways that parallels the English corpus (sentence length, content words, juxtaposed prepositional phrases, two and three juxtaposed PPs); second, it groups together attributes that extended from the English

image-schema (type-tokens and independent constructions); third, it illustrates the attributes in which English translation resembles neither BP nor English (attention units) or in which English translation interacts with both text categories, appearing as a fuzzy category (sequences of four PPs.); fourth it illustrates the mean rank sums of those cases in which the difference was not statistically significant (nominalizations and time and locative adverbials).

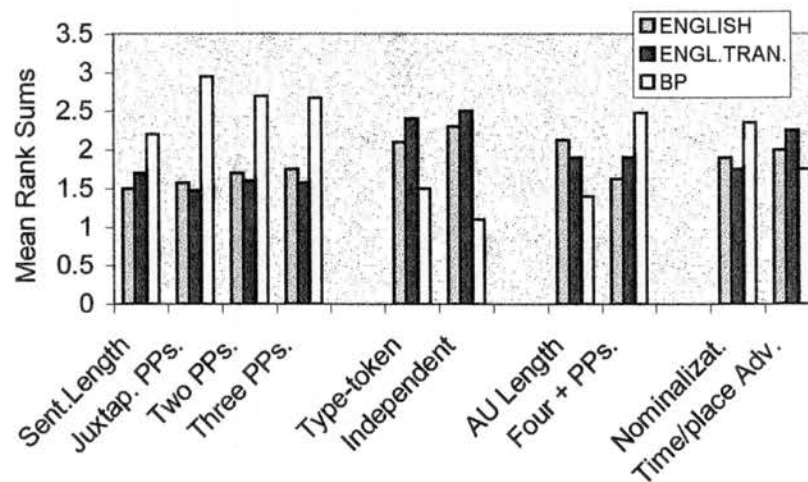


Figure 24. Comparing the mean rank sums in the three research conditions.

Thus, discourse translation appears as a more typical member of the English text category in the way translators use sentence length, content words, juxtaposed PPs, and sequences of two, and three PPs. Yet, in the case of the type-token ratio and number of independent sentences/clauses per thousand words, though English translation interacts

with English texts statistically, the mean rank sums are higher than those of English. This suggests that for these attributes English translation extends from the typical English text. For attributes such as length of attention units, and sequences of four or more PPs, English translation emerges as a fuzzy category. In the case of attention units it takes distance both from BP and English, and in the case of juxtaposed four or more PPs it is equally close to both categories.

As in the case of BP and English discourse, the quantitative results reflect just part of the story. To have a complete picture of the ways in which the translation variant constitutes a category at the boundaries of the English and BP discourse categories, we need to examine different stretches of discourse in English translation. In doing that, we will be able to address the cognitive, perceptual, and pragmatic factors that lead translators to adjust the organization and use of the selected attributes.

Qualitative analysis

The text below comes from *Oil production in the Amazon* (Petrobras Public Affairs Service, 1994). A comparison between the discourse in BP and English translation s illustrates how the translator amalgamated cognitive and pragmatic considerations, including assumptions about what needs to be communicated, in one coherent text. Note that though both texts consist of two sentences, the BP text shows lengthier sentences than its English variant, as well as combines more information in fewer words:

O campo de Rio Urucu, a 600 quilômetros
 the (m.) field (m) of river Urucu **which is** 600 kilometers

de Manaus, concretizou, depois de mais de sete
 from Manaus, made true (3rd.p.s; past) after more than seven

déadas de pesquisas, o sonho de produzir petróleo
 decades of research (pl.,s., f.) the (m,sg) dream (m,s) of producing oil (m,s)

na Amazônia. E um óleo leve, excelente
 in the Amazon. It is (3rd.p.s.) an oil light excellent

para a produção de derivados nobres,
 for the production of by-products (m,s.) high-grade

transportado em longas viagens
 it is carried (m,s.) in long (f,pl.) trips (f,pl.)

pelos rios até a Refinaria de Manaus, onde se transforma
 by (m,pl.) rivers (m,pl.) until the (f,s.) refinery (f,s.) of Manaus **where it is transformed**
 (3rd.p.s.)

em gasolina, diesel e outros produtos.
 in gasoline diesel and other (m,pl) products (m,pl).

(1) After over seven decades of prospecting, Brazil's dream of finding oil in the Amazon finally came true. (2) The Rio Urucu field--600 kilometers from Manaus--produces a **light oil /excellent for transforming into high-grade byproducts;** it is carried by river to the Manaus Refinery/ where it is processed into gasoline, diesel and other products.

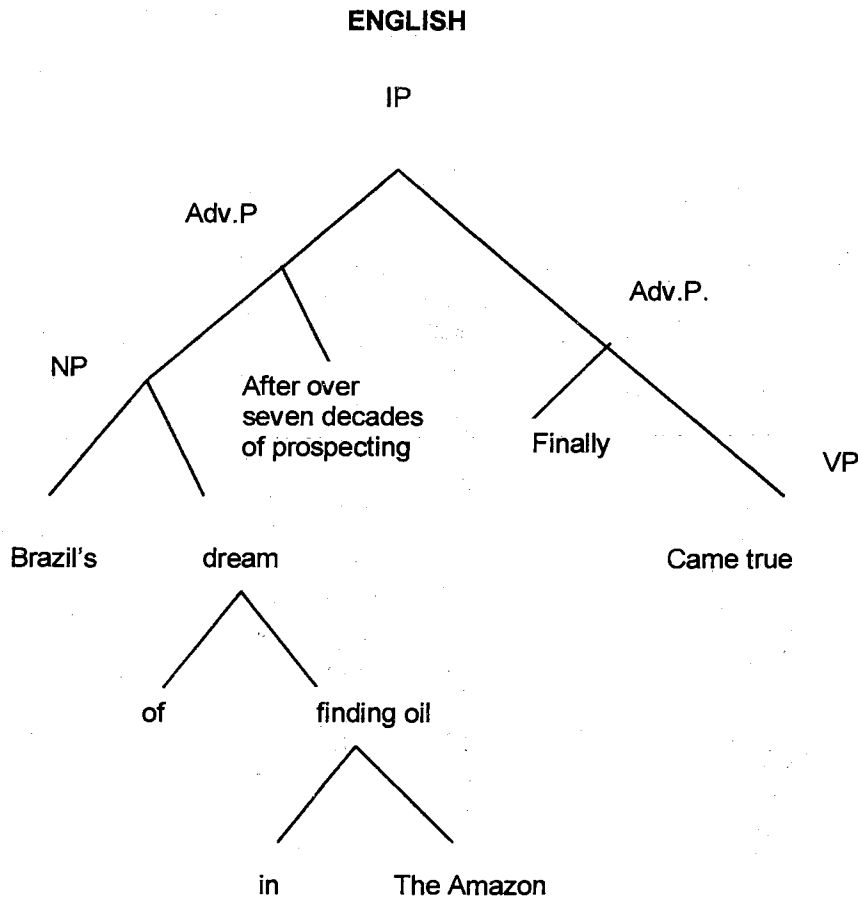
Figure 25 illustrates sentence construction in BP and English translation. The English sentence is sentence (1) above while the BP sentence is the original sentence from which (1) and (2) derived. The figure is particularly efficient to demonstrate that, though English and BP draw on the same constituent structures (NPs, VPs, Adv.Ps, PPs), they organize and use these constituents in sentences in striking different ways. To begin with, the English sentence is just part of the source sentence in the BP text. Then, contrary to the BP text, the English variant shows an adverbial phrase in the first place, immediately followed by the NP --"Brazil's dream of finding oil." This NP functions as foreground information in both texts; however, in BP it is preceded by a long sequence of constituents (see Figure 25), including the introductory Adv.P. of the English variant.

That Adv.P. intrudes between the verb and the final NP “*o sonho de produzir petróleo na Amazônia*” (=the dream of producing oil in the Amazon) in Portuguese.

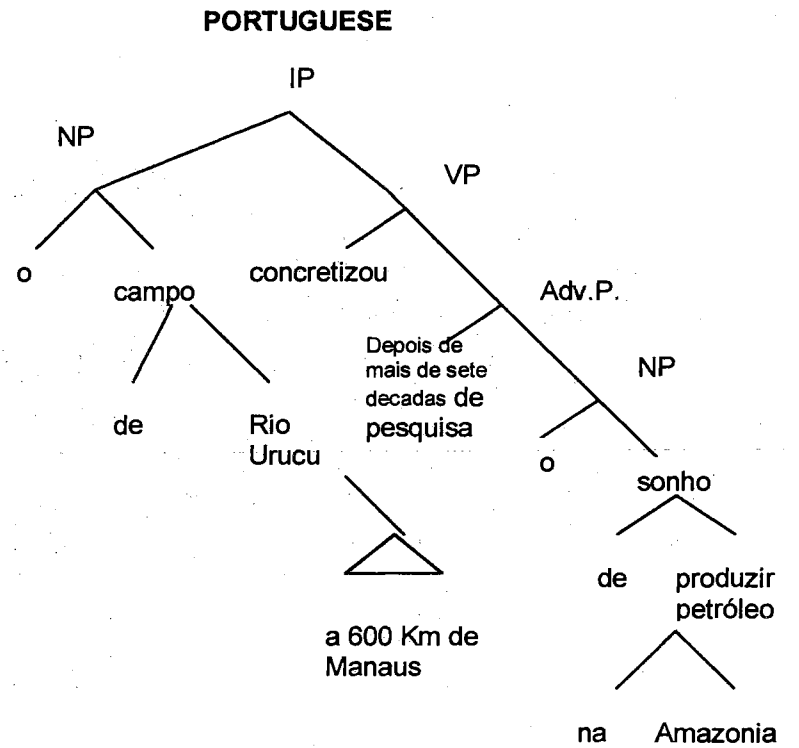
Nevertheless, BP users can keep track of the focus because the NP’s “*o campo*” (=the field) and “*o sonho*” are perceptually more salient in Portuguese than in English translation. They are figures or autonomous entities. In BP, if you recall (see my discussion *Facts about Portuguese*, Chapter V), heads appear first, allowing immediate elaboration of trajectors. In English, on the other hand, modifiers appear first, delaying the elaboration of their trajectors.

Another fact to consider is that the English variant could have placed the NP *Brazil’s dream of finding oil* in sentence initial position moving the time Adv.P. *after over seven decades of prospecting* to sentence final position. However, it thematizes the Adv.P. By using that discourse strategy, the translator activates the attitudinal mental space in which readers should construe the events and the relationships among discourse participants. That is, the Adv.P. signals perspective. It orients the readers about the relevance of what follows. We could even say that it signals that what follows is the climax of years of work, and, perceptually, it functions as the abstract attitudinal setting on top of which discourse develops. This interpretation is reinforced lexically by the use of content words such as “*dream finally came true.*”

The BP readership, contrastively, are members of the speech community in which those facts occurred and know about the difficulties that led to the discovery of the Urucu field through the media at large. If we take in consideration that the GOAL of written discourse is to communicate a message at a minimum cost of processing, the explicit inclusion of ORIENTATION in the BP text would violate the maxim of quantity.



After over seven decades of prospecting, Brazil's dream of finding oil in the Amazon finally came true.



O campo de Rio Urucu, a 600 Km de Manaus, concretizou, depois de mais de sete décadas de pesquisas, o sonho de produzir petróleo na Amazônia.

Figure 25. Comparing the organization of syntax in BP and English translation.

Furthermore, though BP also draws on lexical density to convey images, it does so less frequently than English translation as the quantitative results indicated. In this specific case, instead of reinforcing the implicit discourse orientation through the use of four content forms, BP uses only two: “*sonho*” (=dream) and “*concretizou*” (came true; 3rd.p.sg.). Actually, it lexicalizes the expression “*came true*” in the form of the verb “*concretizou*.” Lexicalization processes and implicit orientation may explain why BP has lower lexical density than English translation.

Note that the word *Brazil*, a landmark, is not present in the source text. But because it orients the international audience about the spatial domain in which the text should be interpreted, the translator inserts it there. Such information was irrelevant in BP because the Brazilian audience is knowledgeable about and closer to the spatial domain in question. It is the translator’s knowledge about the speech event and his sensitivity to grammatical and pragmatic considerations that causes these adjustments. Actually, what the translator tries to do is to map form onto function and thus create a more prototypical extension of the category COMMUNICATIVE TEXT. In doing so, s/he fulfills the GOAL of the speech event. That explains why we find more Adv.Ps. in English translation than in English texts themselves or BP texts.

Figure 25 also illustrates the differences in length of sentences and presence of right branching constructions between BP and English variants. In the diagrammed sentence, the BP variant is lengthier (n= 26) than the English translation (n=17). While BP makes use of right branching (one embedded clause, two Adv.Ps.), English translation does not. Rather, English breaks up the source sentence in two.

The same example also reveals how the English translation moves closer to English COMMUNICATIVE TEXTS in terms of units of attention. While in the source text (BP) the stretches of language between punctuation marks reflect clause-like units or the constituent structure, in the English variant it does it less frequently:

ATTENTION UNITS	WORDS
1. O campo de Rio Urucu,	5
2. a 600 quilômetros de Manaus,	5
3. concretizou,	1
4. depois de mais de sete décadas de pesquisas,	8
5. o sonho de produzir petróleo na Amazônia.	7
6. É um óleo leve,	4
7. excelente para a produção de derivados nobres,	7
8. transportado em longas viagens pelos rios até a Refinaria de Manaus,	11
9. onde se transforma em gasolina, diesel e outros produtos.	9

versus

1. After over seven decades of prospecting,	6
2. Brazil's dream of finding oil in the Amazon finally came true.	11
3. The Rio Urucu field--	4
4. 600 kilometers from Manaus—	4
5. produces a light oil excellent for transforming into high-grade by-products;	12
6. it is carried by river to the Manaus Refinery where it is processed into gasoline,	15
7. diesel and other products.	4

Furthermore, a close look at the word for word translation (see expressions in bold below) of the selected piece of discourse will allow you to see that in five instances one word in Portuguese corresponds to more than one word in English, and in one instance to a whole NP-VP sequence (*transportado* = it is carried). These differences occur because BP rich morphology allows a more flexible word order. Particularly, it allows clauses

without or with ellided subjects. Note that the “dummy” subject “it” appears two times in the English variant while in Portuguese, a series of coordinated right branching constructions, similar to those in Figure 25, appear. These right branching constructions are linked to the trajector by morphology: *petróleo* (s., m.) - *transportado* (s.,m.) - *se transforma* (clitic; s.,m.) and follow a V(s)N (N) pattern:

(S) E um	óleo	leve	/ (S) excelente	para a producao de derivados nobres/
	TR	LM	/	LM
	V	N	/	ADJ (rel.clause)
3rd.p.s.	s.m.	s.		s.

(S) transportado	em longas viagens / onde (S) se transforma	em gasolina
LM(TR)		LM (TR)(rel.. clause)
ADJ (rel clause)		V
s.m.		N
		3rd.p.s.

(note that the landmarks are anaphorically related to the trajector by nominal inflection)

In brief, the variant of the English COMMUNICATIVE TEXT category, which we have just examined, moves away from the BP image-schematic model toward the English image-schematic model because of cognitive, perceptual, and pragmatic considerations. As I hypothesized, the English variant shows shorter sentences and explicit subjects to help readers cope with late closure processing at the constituent and sentence levels, adhering strictly to the (Adv.P.) SV(O) order.

To further examine how those factors operate in the English translation corpus of this study, let's examine an excerpt from Environment, Quality and Safety (Petrobras Public Affairs Office, 1994):

As equipes exploratórias passam vários dias
 The(f.pl.) crews (pl.) exploratory (f.pl.) spend (pl.) several days

em alto mar, nos áridos ... sertões
 in high sea in the (m.) arid (pl.,m.) sertoes (pl.,m.)

ou mesmo nas densas florestas,
 or even in the (pl, f) dense (pl, f) forests

expostas a ambientes hostís em regiões de difícil acesso,
 exposed (pl, f) to environments inhospitable in regions of difficult access

procurando fazer com que sua atividade nômade e efêmera
 trying to assure that their activity nomadic and transient

seja imperceptível aos santuários de vida selvagem.
 be (subj) imperceptible for the sanctuaries of life wild.

(1) **Petrobras exploration crews** spend long periods out in the ocean, in the **Northeast's arid sertao regions**, and even in **Brazil's dense tropical forests**. (2) Exposed to inhospitable environments in remote places, **these teams** do their best to see that their **nomadic, transient activity** will leave no mark on any of **wildlife's havens**.

Once more, the attributes present in BP and in the English variant confirm my expectations and the quantitative results. While BP uses a single sentence (n=42 words) to construe the image of *Exploration in tune with nature*, English requires two (S1=23 words; S2=28 words) for reasons we have discussed previously. That is, BP employs a figure-ground (or head-modifier) pattern and cues anaphoric reference through morphology. By doing that, it facilitates the on-line assignment of semantic roles to the trajectors. Had the Portuguese one-sentence pattern been maintained in the English translation, the English readership would hardly be able to cope with working memory constraints and would probably fail to establish the correct relationship between dependent and autonomous entities. Next, I provide the details of this analysis.

In the English translation version, we have two sentences connected by anaphoric reference to the actor—*Petrobras exploration crews* (first sentence) and “these teams” (second sentence). In BP, however, extensive morphological markings allow the on-line interpretation of head-modifier relations (figure-ground) and reference to the main actor. The words *expostas* (=exposed, feminine, plural) and *sua* (=of the exploratory crews; personal pronoun, fem., 3rd. person) are linked to *equipes exploratórias* (=exploratory crews, feminine, plural) because of gender and plural suffixes. By the same token, right branching constructions appear in sequence, apparently indicating that they do not overload working memory. As you may see below, BP uses a sequence of three adverbial phrases, one embedded non-finite clause, an Adv.P., and two other embedded clauses right after the main clause:

As equipes exploratórias passam vários dias

em alto-mar, (Adv.P)

nos áridos sertões (Adv.P.)

ou mesmo nas densas florestas (Adv.P)

expostas a ambientes hostís (embedded non-finite clause)

em regiões de difícil acesso (Adv.P.)

procurando fazer (embedded non-finite clause)

com que sua atividade nômade e efêmera seja imperceptível aos santuários de vida selvagem. (embedded finite clause)

The dependent clauses modify the main proposition and only connect to it by noun morphology and contextual clues:

As equipes exploratórias
(f.pl.) (pl.) (f.pl.)

expostas a ambientes hostis
ADJ. f.pl.

procurando fazer
contextual clues/ indicate a way of being of the nominal referent or a permanent activity that the nominal referent develops

com que sua atividade nômade e efêmera seja imperceptível...
Det. f.

In the English variant, on the other hand, morphology and contextual clues cannot yield the same effects. Therefore, though we have a similar first sequence of three Adv.Ps.,

in the ocean
in the Northeast's arid *serão* regions
in Brazil's dense tropical forests

we fail to have the successive embedded clauses. Rather, the translator produced a new independent clause that is preceded by an adverbial non-finite clause and followed by a dependent finite clause:

Exposed to inhospitable environments in remote places,
these teams do their best to see
that **their** nomadic, transient activity will leave no mark on any of wildlife's havens.

To keep track of the nominal referent *Petrobras exploration crews*, a full NP reinstated (=these teams) the source of energy in the second sentence. That second mention allows anaphoric reference throughout the rest of discourse (see the possessive pronoun *their*).

Adjustments such as those I have just illustrated, explain the higher frequency of independent clauses in English translation as well as the higher number of content words when compared to the BP source text. The translator is attempting to make the variant a more central member of the English COMMUNICATIVE TEXT family.

In addition, the translator inserted the word *Petrobras*, the main topic entity, in the first NP—*Petrobras exploration crews*—of the English variant, as well as *Northeast*. Those insertions aim at making the spatial domain in which the action took place perceptually more salient to the international audience. As I have pointed out in other instances, that information is part of the background knowledge of the Brazilian audience and, therefore, its explicit mention would violate the maxim of quantity and The Principle of Relevance.

My previous comments about attention units and their relation to clause or constituent structure holds both to the source and the translated text. Actually, the English variant reveals attention units systematically shorter than those of typical English texts, being more BP-like attention units:

Petrobras exploration crews spend long periods out in the ocean,
in the **Northeast's arid sertão regions**,
and even in **Brazil's dense tropical forests**.
Exposed to inhospitable environments in remote places,
these teams do their best to see that their **nomadic**,
transient activity will leave no mark on any of wildlife's havens.

As the Friedman test has demonstrated, the length of attention units in English and English translation differs significantly. In many instances in the English translation corpus attention units seem to be a sub-category of the BP source text rather than of the English COMMUNICATIVE TEXT category.

Observe one more piece of discourse. This excerpt immediately follows the one we have just analyzed above:

A presença da Petrobras na prospecção de petróleo
The presence (f. of Petrobras in the (f.) prospection of oil

nas bacias sedimentares brasileiras
in the (f.pl.) basins (f. pl.) sedimentary (pl.) Brazilian (f. pl.)

é monitorada através de convênios com universidades
is monitored (f.) by agreements (pl.) with universities

e instituições científicas ... em vários estados
and institutions scientific (f., pl.) in various states

garantindo, desta forma, a obtenção de subsídios
guaranteeing this way the capture of subsidies

para que a prospecção de óleo e gás natural
for the (f.) prospection (f) of oil and gas natural

não interfira no ambiente natural.
not interfere (subj.; 3rd. p.s.) in the (m.) environment natural.

(1) Exploratory work in Brazilian sedimentary basin is monitored by agreements with scientific institutions. (2) **This** allows Petrobras to carry out environmental projects **and** interact harmoniously with nature. (3) **In the Amazon, the Universidade Federal do Amazonas and the Federal Environmental Agencies** are helping to research ways of preventing oil and natural-gas exploration from interfering with **one of Brazil's most valuable natural resources--the rain forest.**

Here while Portuguese uses a single sentence to express ideas, the English translation uses three! As in our previous examples, verb and noun morphology, helped by head-modifier relations, disambiguate cues in Portuguese. English translation, in contrast, requires the presence of subjects in clause initial position because its verbs

signal semantic information less frequently. We see then *this* as subject of the second sentence and the insertion of an explicit entity as actor in the third sentence—*The Universidade Federal do Amazonas* and the *Federal Environmental Agency*. These entities are implied (it is hearer/reader conscious) in the Portuguese version. Likewise, in Portuguese there is no such mention as *rain forest* or *most valuable natural resources*. It is up to the BP reader to infer those entities. That is possible for two reasons. First, the entities are available in the AMAZON domain. Second, a photograph of a rig in the Amazon rain forest allows the Brazilian readership to activate the domain and strengthen related implicatures which are part of their experiences.

There is also a repetition of the environmental theme in the English translation: one general (environmental projects) and one specific (Universidade Federal do Amazonas and the Federal Environmental Protection Agency). In BP, however, only a general mention (*ambiente natural* = natural environment) appears. The translator, by explicitly including the specific mention takes the English readership one step further into the concrete reality of the events in the discourse. Furthermore, there might also have been a slight different ORIENTATION given the GOAL of the speech event. While it is necessary to sell an image of sound familiarity with environmental protection measures to the international audience, that may not be true for the Brazilian audience.

Consider also the following piece of discourse extracted from *Transporting Safely* (Environment, Quality and Safety, Petrobras, 1994):

<i>Os</i>	<i>dutos,</i>	<i>além de</i>	<i>passarem</i>	<i>por</i>	<i>inspeções</i>	<i>freqüentes,</i>
The (m.pl.)	pipelines,	besides	going (3rd.p. pl.)	through	inspections	frequent (pl.),
<i>São</i>	<i>dotados</i>	<i>de</i>	<i>dispositivos</i>	<i>de</i>	<i>segurança,</i>	<i>como</i>
are (3rd.p.pl.)	equipped (m.pl.)	with	devices (m.pl.)	of	safety	such as

signal semantic information less frequently. We see then *this* as subject of the second sentence and the insertion of an explicit entity as actor in the third sentence—*The Universidade Federal do Amazonas* and the *Federal Environmental Agency*. These entities are implied (it is hearer/reader conscious) in the Portuguese version. Likewise, in Portuguese there is no such mention as *rain forest* or *most valuable natural resources*. It is up to the BP reader to infer those entities. That is possible for two reasons. First, the entities are available in the AMAZON domain. Second, a photograph of a rig in the Amazon rain forest allows the Brazilian readership to activate the domain and strengthen related implicatures which are part of their experiences.

There is also a repetition of the environmental theme in the English translation: one general (environmental projects) and one specific (Universidade Federal do Amazonas and the Federal Environmental Protection Agency). In BP, however, only a general mention (*ambiente natural* = natural environment) appears. The translator, by explicitly including the specific mention takes the English readership one step further into the concrete reality of the events in the discourse. Furthermore, there might also have been a slight different ORIENTATION given the GOAL of the speech event. While it is necessary to sell an image of sound familiarity with environmental protection measures to the international audience, that may not be true for the Brazilian audience.

Consider also the following piece of discourse extracted from *Transporting Safely* (Environment, Quality and Safety, Petrobras, 1994):

<i>Os</i>	<i>dutos,</i>	<i>além de</i>	<i>passarem</i>	<i>por</i>	<i>inspeções</i>	<i>freqüentes,</i>
The (m.pl.)	pipelines,	besides	going (3rd.p. pl.)	through	inspections	frequent (pl.),
<i>São</i>	<i>dotados</i>	<i>de</i>	<i>dispositivos</i>	<i>de</i>	<i>segurança,</i>	<i>como</i>
are (3rd.p.pl.)	equipped (m.pl.)	with	devices (m.pl.)	of	safety	such as

válvulas de bloqueio, que impedem a passagem
 valves (f.pl.) of shut-off, that prevent the escape

de produtos em caso de anormalidade,
 of products in case of abnormalities,

resguardando as condições naturais das áreas marginais.
 Protecting the (f.pl.) conditions (pl.) natural (pl.) of the (f.pl.) areas (pl.) surrounding (pl.).

(1) Petrobras routinely inspects its **pipelines**. **These facilities** are equipped with such safety devices as **shut-off valves, which prevent** inadvertent escape of products in case of abnormalities and thus **protect** the natural conditions of surrounding areas.

Again, what was a single sentence in BP becomes two sentences in the English translation. Because of the reduced appositive clause (*além de passarem por inspeções freqüentes*) in between the subject-NP (*Os dutos*) and the VP (*são dotados de dispositivos de segurança*) and the sequence of a relative clause (*que impedem a passagem de produtos em caso de anormalidade*) and a reduced gerundial clause (*resguardando as condições naturais das áreas marginais*), that sentence was particularly prone to ambiguity in English. Therefore, the translator profiled the appositive clause *além de passarem por inspeções freqüentes* (=Besides going through frequent inspections) as an argument. S/He transformed the appositive clause into a main participant of a typical action chain where Petrobras appears as the source of energy. That explicit actor does not figure in the BP text. Rather, it is reader conscious.

Moreover, the translator also profiled the verb as a typical TEMPORAL PROCESS (=Petrobras inspects...). Those two strategies created room for cohesive ties (anaphoric reference, lexical and syntactic) that avoid any potential competition of cues during discourse comprehension. That is, “pipelines” (=dutos) appear as the argument of the

first sentence and re-appears, in a second mention, as the topic of the second sentence (=These facilities). The argument of the second sentence (=safety devices) becomes the topic and actor of the next two clauses by means of the relative pronoun “which.” See the expressions in bold in the English translation of the excerpt.

By breaking the sentence in two and including a second mention of *pipelines*, the translator could also use the resources English syntax offers to clarify the relation between the trajector *shut-off valves* and the landmarks that modify it *which prevent...and thus protect.....*. Cohesive ties such as parallelism and consistency of verb form collude to build a clear image of the passage as a whole.

In the source text, though, there is no explicit mention of an actor. Contextual effects generated by linguistic information in the previous discourse clarify any potential ambiguity. In the very first paragraph of the text, *Petrobras* is set up as the company in charge of oil distribution in Brazil:

<i>Para</i>	<i>que</i>	<i>o</i>	<i>petróleo</i>	<i>..</i>	<i>chegue</i>	<i>às</i>	<i>refinarias</i>
In order	to	the (m.)	oil (m)		get	to the (f.pl.)	refineries (f.pl.)
<i>e</i>	<i>os</i>	<i>derivados</i>	<i>.</i>	<i>ao</i>	<i>consumidor,</i>		
and	the (m.pl.)	oil products (m.pl.)		to the (m.)	consumer (m.)		
<i>a</i>	<i>. Petrobras</i>	<i>opera</i>	<i>extensa</i>	<i>rede</i>	<i>dutoviária</i>		
the (f.)	Petrobras	operates	extensive	network	pipeline		

It remains reader conscious or semi-active throughout the discourse. Had it been mentioned a second time, it would cause an “unjustified or gratuitous” processing effort (Sperber and Wilson, 1995). Therefore, there is not a second explicit or even implicit

mention of the actor. Instead, passive constructions give cognitive prominence to Petrobras' actions, changing the ORIENTATION of discourse:

São utilizados navios petroleiros...
Are used (m.pl.) tankers (m.pl.)

Os terminais marítimos estão dotados de equipamentos...
The (m.pl.) terminals (m.pl.) maritime (m.pl.) are equipped (m.pl.) with equipments

Centros de combate à poluição foram instalados...
Centers (m.pl.) of control to pollution were installed (m.pl.)

Os dutos são dotados de...
The (m.pl.) pipelines (m.pl.) are equipped (m.pl.) with

or the action becomes the actor in the form of a nominalization:

A movimentação de combustíveis exige rigorosas medidas de prevenção
The (f) moving (f) of fuels demands strict (f.pl.) measures (f.pl.) of safety

Also, as in other cases we have discussed in this study, BP morphology (see expressions in bold) appears as the number one guide to comprehension. We have local incremental gains as we move on in discourse, until we cumulate the different stimuli to form the composite structure.

It is worth mentioning as well that the Brazilian readership is expected to know from context and experience that Petrobras is the source of energy. The assumption is communicated as an implicature. It is manifest that the translator considers sub-sequent mentions of *Petrobras* irrelevant. Therefore, s/he maximizes the cumulative relevance of endophoric and exophoric stimuli.

In the English translation, however, there is at least one explicit mention of *Petrobras*, as the actor, in every single paragraph (five mentions in total). Based on

Sperber and Wilson's Relevance Theory (1995), this stimulus is worth the addressees' effort to process it. In addition, the fact that the translator includes those mentions in the English translation variant tells us s/he presumes it is relevant to the readers. The explicit mentions of *Petrobras* prevent them from creating unintended implicatures. That is, the mentions maximize the manifestness of the information writers want readers to have. Furthermore, such discourse strategy allows the translator to use the few resources of English morphology, leading to local cognitive gains .

In brief, the translator, sensitive to cognitive variables which affect discourse construction in English, introduced discourse entities, broke the original BP sentences in shorter sentences, and changed orientation. In doing so, s/he set an abstract mental space against which English readers could map form, created textual structure for further development of discourse, and reduced possible threats to sentence comprehension, mainly ambiguous reference.

I would like to go back to two findings I reported earlier in this study. First, the lower frequency in which BP uses content words in each thousand. Second, the relation between length of attention units, amount of information in each unit, and constituent structure. Take the previous discourse example.

Os dutos,
The (m.pl.) pipelines (m.pl.),

além de passarem por inspeções frequentes,
besides to go (Inf;3rd.p. pl.) through inspections (m.pl.) frequent (pl.)

são dotados de dispositivos de segurança,
are equipped (m.pl.) with devices (m.pl.) of safety

como válvulas de bloqueio,
such as valves of shut-off

que impedem a passagem de produtos em caso the anormalidade,
 that prevent (3rd. p. pl.) the (f) escape of products (pl.) in case of abnormality

resguardando as condições naturais das áreas marginais.
 Protecting (ger.) the (f.pl.) conditions (f.pl) natural (pl.) of the (f.pl.) areas (f.pl.) surrounding (pl.)

- (1) Petrobras routinely inspects its **pipelines**.
- (2) **These facilities** are equipped with such safety devices as **shut-off valves**,
- (3) **which prevent** inadvertent escape of products in case of abnormalities and thus **protect** the natural conditions of surrounding areas.

This piece of English translation discourse constitutes a less peripheral member of the “communicative” English discourse category in terms of number of attention units per sentence, length of attention units, and precise lexical choice. Its three attention units express ideas with five, eleven, and nineteen words respectively, out of which 21 are content words. Also, the three units are lengthier than those in the source text. The source text, on the other hand, displays six attention units with two, six, six, four, ten, and seven words out of which 20 are content words. The difference among English translation and BP units is not only in number, but in the amount of information each unit carries. While in BP most words (content or not) guide processing through morphology (see the indication of the morphological trappings in bold), in English translation, syntax does the job. The English discourse variant uses an independent and two coordinated relative clauses. Therefore, there are fewer but longer units.

Another way of emphasizing the importance of cognitive, perceptual, and linguistic variables in discourse translation is by examining instances in which attributes English users expect to see in “communicative” English texts fail to appear. When that happens,

users are forced to reach back and reactivate the images or relations being profiled.

When the variant construed is a very peripheral extension of the COMMUNICATIVE category “English text,” users get lost and comprehension may fail. I will provide an example from *Environment, quality and safety* (Petrobras, 1994):

Em um cenário mundial cada vez mais competitivo,
In a (m., s.) scenery (m., s.) world each time more competitive (m., s.)

a sobrevivência e a imagem das empresas
the (f., s.) survival (f.,s.) and the (f.,s.) image (f.,s.) of the (f., pl.) companies (f., pl.)

estao diretamente relacionadas à conscientização de seus empregados
are directly related (pl, fm) to the awareness.....of their employees

de que o aprimoramento contínuo da qualidade
that the (s.,m.) refining (m,s.) continuous (m,s.) of (s.,f.) quality

a busca dedicada de sistemas e procedimentos
the (f., s.) search (f., s.) dedicated (f., s.) of systems and procedures

que visem a minimização de impactos ao
that view the (s., f.) minimization (f., s.) of impacts (pl., m.) to the (s., m.)

meio ambiente e a segurança do homem
environment and to the (s., f) safety of the (s., m) human being

e do patrimônio são fatores decisivos
and of the asset (m., s.) are factors decisive (m., pl.)

para o aumento da produtividade e da eficácia.
to the (m., s.) increase of the (f., s.) productivity and of the (f., s.) efficiency.

In an increasingly competitive world, a corporation’s image—and indeed its very survival—is inseparable from **keen staff awareness** // that **steady upgrading of quality, the incessant search for environmentally sound systems and procedures, and the protection of human lives and property** are decisive in boosting productivity and efficiency.

When English readers try to process sentences like this, they reach the end and they do not know what kind of relation or process is being profiled. They have to reach back and reactivate the autonomous entities to which each landmark refers. See the case of *keen staff awareness* in the above example. To make things even worse, within the nominal clause that complements *keen staff awareness*, we have three long sequences of head-modifier constituents:

steady upgrading of quality
the incessant search for environmentally sound systems and procedures
the protection of human lives and property

Before reaching the trajector, readers have to be able keep all landmarks active in working memory. However, it takes such a long time to get there that when they do, they may have to start all over again. The same effect does not occur in the BP variant for two reasons. First, the head-first pattern immediately instantiates a core image. Second, the gender and plural trappings attached to the determiners, prepositions, adjectives and profiled nouns facilitate on-line processing as well. The linguistic devices keep readers focused. Had cognitive, perceptual, and linguistic variables been taken in consideration in the English variant of the discourse, this sentence would perhaps read like:

In an increasingly competitive world, the image and survival of a corporation **depends** on its ability **to show** its employees that productivity and efficiency are functions of a continuous search for quality and for systems and procedures **which** minimally **impact** the environment and human lives.

In my translation of the text, I tried to show first the autonomous entities. They should be perceptually more salient because they set up the mental space on which to

map form. In doing so, I avoided possessive constructions of dubious idiomaticity like *corporation's image*, using instead a sequence of NP PP to bring the autonomous entities closer to the readers. Also, I tried to profile processes instead of states. To do so, I inserted predicates like *depends*, *to show*, and *impact*, foregrounding their arguments and allowing English morphology to signal semantic information. Last, I deleted unnecessary modifier-head sequences like *quality improvement*. If I am in continuous search for quality, I am trying to improve it each day. The resulting discourse is closer to a prototypical "communicative" English text. I have to say that in my experience as a translator NP's N and Adj (Adj) (Adj) N constructions caused the greatest difficulties in English. Because English is a head final language, there is a tendency to believe that anything, no matter how much, can be placed before the head or that any entity, no matter its degree of inanimacy or abstractness can equally well serve as a possessor in possessive genitive constructions. When that occurs, we end up with marginal exemplars of possessive genitive and noun sequence constructions in English. In addition, that may overload the English reader's working memory, offsetting important perceptual cues to meaning. Cognitive variables such as cost of processing, measured as the amount of time it takes to build a composite image structure, and perceptual salience are fundamental considerations to successfully construe more prototypical members of the COMMUNICATIVE TEXT.

Actually, the number and length of head-modifier sequences brings to mind the scarce number of juxtaposed PPs in the English translation corpus. One of the facts that leads this corpus to exhibit fewer juxtaposed PPs and sequences of two and three PPs than typical English discourse is the frequent and successive additions of modifiers

before a head to express what in the source text are head-complement relations. This is basically the case of the previous selection. Another illustration of the frequent and lengthy modifier-head sequences is the example below:

Petrobras, the state-run holding company responsible for integrated oil-industry activities in Brazil, is an internationally recognized master of state-of-the art deepwater exploration and production technology and boasts a world-class technical staff.

In trying to transform BP discourse into a typical English discourse, the translator overindulges in the use of modifier-head sequences. This makes texts similar to the selections move away from the typical English discourse or even away from the English image-schematic model. Though English uses head-modifier sequences because they are part of English structural characteristics, it does not use them so extensively. Rather, it balances it with other constructions that can place autonomous entities in perceptually salient positions. See the translation example I provided.

Thus, contrary to what the quantitative results indicated, the fact that English translation reveals fewer juxtaposed PPs or sequences of two and three PPs than typical English discourse generally makes English translation a fuzzy category. In addition, it forces readers to re-arrange conceptual structures with more frequency, yielding an unjustified processing effort. Many times, it impacts comprehension.

In relation to sequences of four or more PPs, every time that they appear, they replicate the structure present in the source text:

1. ...where small barges loaded the oil for transportation

**to the town
of Coari
on the banks
of the Solimões River**

*Balsas de pequeno porte levavam o óleo
até a cidade
de Coari,
às margens
do Rio Solimões*

2. A new oil pipeline was built

**over the 57 kilometers
between the production area and the River Terminal
on the banks
of the Tefé river.**

*Um novo leoduto foi construído
nos 57 quilômetros
entre a área de produção e o terminal
nas margens
do rio Tefé*

3. This is perhaps why it is known as one of the areas
with the highest incidence
of malaria
in the entire State.

*Talvez por este motivo, é reconhecido como uma
das áreas
de maior incidência
de malária
em todo o estado*

These examples illustrate how English translation resembles BP discourse in its use and organization of four or more juxtaposed PPs. On the other hand, the low frequency in which English translation uses more than four juxtaposed PPs reminds us of English texts. Despite Brazilian Portuguese's constant use of sequences of six, ten, even 13 juxtaposed PPs, English translation exhibits only one instance of 5 juxtaposed PPs in its corpus. These facts corroborate the quantitative results and reinforce the interpretation of English translation as a fuzzy category.

In case you wish to further entertain these ideas, Appendix D displays English and Portuguese versions of three other texts which were part of the data side by side. In the English version, I marked entities that were inserted or deleted as well as every paragraph in which the English text uses more and shorter sentences than the BP text.

Summary. English translation discourse takes distance from BP discourse in the way it uses and organizes attention units, type-tokens, content words, and independent constructions into sentences. On the other hand, English translation resembles other members of the English text category in its way of using sentence length and content words. In addition, the texts in the corpus extend from the typical English COMMUNICATIVE TEXT in the way they use independent constructions and type-tokens. With regard to attention units, juxtaposed PPs, sequences of four PPs, and modifier-head sequences, English translation yielded mixed results. The quantitative analysis suggested that English translation uses juxtaposed PPs in ways similar to English. However, qualitative analysis indicated that English translation's lower frequency of juxtaposed PPs relate to an increase in the number and length of modifier-head sequences that may overload readers' attention resources and impact comprehension. In addition, qualitative

analysis demonstrated that English translation attention units were systematically shorter than those of typical English texts, being more BP-like attention units. In brief, English translation emerged as a fuzzy category, or a transformation of BP discourse toward the image-schematic model of English discourse. It resembles English discourse in many instances, but in some others, it still reveals characteristics typical of BP discourse. Finally, it was demonstrated that translators need to consider cognitive, perceptual, and cognitive factors to produce texts that allow optimal processing in the target language.

Summary of findings

In this dissertation, I examined how Brazilian Portuguese and English discourses organize and use different attributes along the syntactic, lexical, and reality-interaction axis. I also tried to explain the different ways in which the attributes lead to hypothesis about conceptual organization in the two languages. Linguistic attributes interact with cognitive, perceptual, pragmatic, and rhetoric variables to construe the PATH gestalt of a COMMUNICATIVE TEXT image-schema.

Some of the attributes of the English COMMUNICATIVE TEXT include short sentences; long attention units; a combination of independent and non-finite clauses; few embedded clauses; a high number of nominalizations; explicit actors; precise choice of collocated content types; explicit interaction with readers' reality; reliance on the lexicon and pronouns to keep the referents active or semi-active. Some of the attributes of the BP COMMUNICATIVE TEXT include long sentences; attention units that generally reflect clause or constituent structure; frequent embedding and right branching constructions, specially by means of juxtaposed PPs and non-finite clauses; absence of explicit actors; implicit interaction with readers' reality; perceptual salience of autonomous entities; few

nominalizations; frequent repetition of content words that cue the topic; reliance on morphology to keep autonomous entities active or semi-active (see Figures 19 and 20 in this chapter).

The co-occurrence of these attributes constitutes cases of “English texts” or “Portuguese texts” close to the prototypical. The absence of one or more of these attributes constitutes instances of variation within the category or less prototypical cases. English translation is one such example. It appears as a member of the English text family in its use of sentence length, content words, type-tokens, and independent constructions. Still, in other instances, it appears at the boundaries of English and BP image-schematic models. It not only overuses modifier-head sequences in a non-English way but also frequently uses BP-like attention units and sequences of four PPs.

Discussion

Languages seem to balance the way in which they distribute linguistic devices in written discourse, aiming at optimal comprehension. Sentence length, for example, appears to be a function of the cognitive effects yielded by linguistic attributes typical of different languages. The COMMUNICATIVE BP institutional written discourse uses sentences of a median length of 24 words. This is more than human attention resources can keep active in the processing of discourse (see chapter III). To counterbalance sentence length, BP typically uses a high number of brief attention units (median=7) that resemble constituent or clause-like structures and that can be managed by working attention according to existing research. In addition, BP institutional discourse tends to repeat words that have already appeared in the text and to establish correspondences among events by means of morphologic trappings. Both repetitions and morphologic trappings function as cohesive ties and facilitate the allocation of attention resources.

Overt morphologic and lexical ties allow BP to chain right-branching constructions such as non-finite clauses and sequences of prepositional phrases generally without ambiguity. These dependent constructions elaborate the moderate number of independent clauses/sentences (40:1000) or figures that appear in discourse and relate to them, as well as to other right branching constructions, through morphology. Right branching constructions in BP also relate to the autonomous entity and to other right branching constructions through collocation with the topic of discourse.

Another characteristic of BP discourse that seems to counterbalance sentence length is word order. Head-modifier relations and clauses tend to appear in the order they will be conceptualized. That is, trajectors or autonomous entities appear first and landmarks appear second. According to Cognitive Grammar, this means that entities are activated and conceptualized in the order they appear. In other words, Brazilian Portuguese word order favors on-line processing of written information.

The idea that word order in BP favors on-line processing of written information is reinforced by the network of morphologic and lexical ties. It is also reinforced by the number of attention units that parallel constituent and clause structure. If this is true, sentence length offers little constraint to production and comprehension processes in BP. Conceptualizers activate and layer words in the order they appear, keep referents and topics active by means of morphology and repetition, and manage attention resources by means of units that stay most of the time within the limits of working memory capacity.

On the other hand, the COMMUNICATIVE English institutional written discourse uses more, but significantly shorter sentences of a median length of 20 words if compared to BP institutional discourse. The presence of sentences that are shorter than BP sentences counterbalances the fewer morphologic resources English uses to keep

reference active. It also compensates for word order that leads users to rearrange component structures when construing head-modifier composite schematic structures. That is, short sentences compensate for the need to delay the construal of final composite structures.

Another characteristic of English discourse that seems to counterbalance word order effects and few morphologic resources is the number of independent constructions. English writers seem to balance the number of independent constructions (53:1000) with that of right branching constructions with a slight tendency to rely more on independent constructions. Also, though they use non-finite clauses and sequences of juxtaposed PPs to develop topics, they use them with a significant lower frequency than BP writers. Actually, sequences of three juxtaposed PPs were rare in the English corpus, and sequences of four or more PPs nearly non-existing. This is congruent with the fact that English has few morphologic resources to act as overt cohesive ties. When English writers use right branching constructions, they cannot tie them to the autonomous entities by means of morphology. In addition, pronominal reference many times fails to work and gives rise to ambiguous constructions. This phenomenon has been well illustrated in the study of translated discourse. Sentences in BP with extensive right branching or with embedded clauses were broken in two or three sentences in the English translation variant to keep autonomous entities in a perceptually salient position and avoid ambiguity. The shorter the sentence, the less ground and the more figure participants it tends to have. This seems to be the case of sentences in English.

Another characteristic of English discourse that contributes to its short sentences is the median length of attention units. Though a nine-word attention unit complies with the limitations of working memory, it does not facilitate management of attention

resources as much as BP shorter attention units. Besides, multi-clausal attention units also appeared with frequency in the English corpus. They definitely require more effort from working attention than BP short units. Thus, sentence length in English also seems to be a function of the amount of information encoded in attention units.

Sentence length in English also counterbalances the high type-token ratio and the number of collocations qualitative analysis brought to light. To weave “texture” and establish correspondences among frequently used autonomous entities such as nominalizations (61:1000) and independent clauses, English apparently relies on the lexicon. If readers are not familiar with some of the words writers use, varied vocabulary and frequent collocations may also impact efficient management of working attention. Short sentences then become an aid toward comprehension. They generally help readers keep track of the network of words.

In brief, it seems that short sentences in English make readers more capable of coping with any difficulty that might arise from precise lexical choices, rarer repetitions, a more varied vocabulary, frequent nominalizations, few morphologic trappings that could function as anaphors, and a word order that leads to mental rearrangement and, ultimately, to a late-closure processing strategy.

The fact that BP (65:1000) and English (61:1000) used nominalizations with equal frequency reinforces the interpretation that languages balance linguistic events toward optimal comprehension. Nominalizations are clause-like constructions and the more nominalizations texts have, the more time consuming it would be to process sentences with a median length of 24 words. To counterbalance the number of nominalizations, BP relies on morphology, on the situation context, and on collocation with the topic. The cases of nominalizations analyzed in the qualitative analysis ended up yielding positive

cognitive effects. English, on the other hand, counterbalances the number of nominalizations by using sentences with a median length of 20 words, a fact that apparently opens space for processing more nominalizations. Also, nominalizations are typically cohesive ties between sentences. If there are more sentences in English, we would expect frequent overt cohesive ties.

In addition, facts related to lexical variety and lexical density in BP are congruent with the fact that BP encodes different grammatical functions in one token and has grammaticized other concepts in one token. For example, notions of aspect, tense, mood and number appear in one form in BP. The preposition *em* (=in) contracts with the article *o* (=the) and its inflections forming single forms. The same is true of the preposition *de* (=of). Besides, morphologic trappings lead users to keep nominal reference under the focus of attention and make the reinstatement of full NPs, nouns or pronouns unjustified.

In many instances that English had to reinstate full NPs or use a pronoun, BP simply carried the notion in the morphological trappings. Had BP writers reinstated the NP or used the pronoun, they would be violating The Principles of Relevance, or else, they would be forcing readers to go through unjustified processing effort. Thus, the lower type-token ratio and lexical density in BP are functions of morphology and The Principles of Relevance.

Verbs at clause initial position and no agents are also a function of morphology. While conducting the qualitative analysis for this study, I noticed that BP tends to delete agents and keep verbs at clause initial position with frequency. The agents appear in the verb morphology and are semi-active in readers' working attention. Also, similarly to Verhagen's (1996) findings about extraposed relative clauses and the verbs *stehen* and *liegen* in German, verbs in BP may frequently appear first to indicate that they should be

conceptualized independently and construed objectively. This, however, is a topic for another dissertation.

In other words, the type-token ratio as well as the number of content words in BP derive from grammaticization processes and verb morphology. It is no wonder that the BP type-token ratio and lexical density are lower in the BP corpus than in the English corpus. Lexically, what really seems to distinguish the two research conditions is how each uses those tokens to cue meaning. One relies on repetitions and the other on collocations. The two strategies seem congruent with the fact that BP uses longer sentences than English and therefore needs to ease the effort of working attention.

The texts I analyzed provide evidence for the above conclusions as much as other studies involving direct manipulation of processing strategies in languages closely related to Portuguese, namely Italian, Spanish, and French (see the review on sentence processing in Chapter IV). Language specific morphology and word order pose cognitive constraints on sentence production and processing, as extensively demonstrated by previous crosslinguistic research (Bates and MacWhinney, 1989 among others; see chapter IV).

In other words, different languages have different properties that make texts more or less prototypical exemplars of what I called a COMMUNICATIVE TEXT image-schema. A central exemplar of the COMMUNICATIVE TEXT schema yields maximum cognitive effects, being cognitively more efficient than less prototypical exemplars. It is the unmarked instantiation of the schema. The one that gives the schema its shape, is easier to use and remember, and motivates variations within the category by interaction with other schemas and with the conceptualizer's own model of texts.

Next, I discuss the results of this study in view of the literature review (Chapters II, III, and IV) and the major issues it brought up. Namely, I will address the relation between attention units and working attention, differences between speech and writing, genre-related conclusions, linguistic relativity and the psycho-pragmatic approach, social and cultural phenomena, and the translatability of conceptual structures.

Results in the light of the literature review

BP discourse features in the light of existing studies. Findings of this study corroborate Oliveira (1997a; 1997b) and provide strong evidence for the more elaborated style of expository BP discourse. Expository prose in BP has fewer and lengthier sentences and makes more frequent use of subordination in complex sentences. The roots of this more elaborated style may go back, in part, to a European literacy tradition, as Oliveira indicates. On the other hand, it may also arise from what grammar allows users to construe in view of the allocation of attention resources. Given that those very literacy traditions stem from languages that are structurally similar to Brazilian Portuguese, grammar seems to play a considerable role.

With regard to contextual explicitness, this study contradicts Oliveira (1997b) in the sense that texts in BP did not mark, explicitly, geographic and time contexts. Actually, the qualitative analysis indicated that BP expository discourse seems to be less explicit than English in that it uses fewer explicit mentions about the situation context. However, the qualitative analysis has also demonstrated that, quite frequently, BP maximizes cues present in the situation context to make the same idea salient. In other words, the context is implicated by cues in the text and by the situational model. The different results for the contextual explicitness dimension may also derive from the attributes included in this factor. In addition to geographic references, Oliveira (1997a) has counted cultural,

historic, politic, and socio-economic references. As she correctly points out, because of Brazil's distinct social, economic, and political situation, students may have been led to mark more intensively references to the context to build a scenario. Other qualitative studies could perhaps better clarify this point.

Furthermore, findings of this study reinforce the thorny methodological difficulty of selecting and counting attributes that are equivalent in the languages contrasted (Lux and Grabe, 1991; Biber, 1995; Connor, 1996; Oliveira, 1997b). Lux and Grabe bring to light the problematic counting of personal pronouns and prepositions when contrasting BP to English. While personal pronouns may be deleted in BP, in English they need to be explicit. Furthermore, as opposed to English, Portuguese draws extensively on prepositional phrases because it is a head-final language. Verb and noun sub-categorization appears with high frequency. This is why I eliminated the counting of PPs that derived from verb sub-categorization from this study. This is why I only tallied sequences of two or more juxtaposed PPs. This is also why I don't believe personal pronouns should appear in the informational-interaction axis if Portuguese is one of the languages in the comparison. If morphology refers to nominals and personal pronouns, the inclusion of personal pronouns in BP discourse most of the time represent an unjustified cognitive cost for readers. Their inclusion in the informational-interactional axis blurs results.

Further elaborating on the methodological difficulty of selecting equivalent attributes, Oliveira (1997a) highlights the different ways BP and English mark cause-effect relations. BP expresses them by means of nouns, verbs, conjunctions, prepositions or locutions, not only by means of conditional or concession clauses introduced by the connective *if*, or reason clauses that convey a direct relationship with the main clause.

She cautions that contrastive studies that adopt the MD methodology should investigate if the attributes are functionally equivalent in the languages contrasted to avoid blurred results. Most of all, she calls attention to the fact that these studies, including hers, adopt functional descriptions that are true of English given that they have not been appropriately investigated in BP.

I not only endorse her words but also raise questions about the inclusion of repetitions in the informational-interactional axis by studies that used Biber's (1988) Multidimensional Model and that addressed BP discourse (e.g. Dantas-Whitney and Grabe 1989; Oliveira, 1997b). Repetitions, as I have mentioned, ease the allocation of attention resources and counterbalance sentence length. Function seems to be interacting with the cognitive role of linguistic devices to determine the kind of attribute to use and the frequency of use.

Therefore, following The Principles of Relevance, I do not believe the presence or absence of pronouns, full NPs or repetitions in BP could be related to the informational-interactional axis. Rather, I suggest that they relate to cost of processing. As mentioned before, because of morphology, BP will generally show fewer pronouns and full NPs than English will. It will also use repetitions with greater frequency. Studies that adopt Biber's (1988) model should perhaps investigate not only the form-function relation of attributes selected for analysis, but also, their psycho-pragmatic roles.

In addition, my results depart from Dantas-Whitney and Grabe (1989). While this study indicates that institutional expository prose in English and BP is informational-oriented, Dantas-Whitney and Grabe indicated that BP editorials typically used more concrete and colloquial style than English did. It is true that those differences may relate to genre idiosyncrasies given that Oliveira (1997b) found a more formal and

informational style to be typical of BP freshmen compositions. However, in the light of this dissertation and of Oliveira, I suspect the factors in the dimensions of Dantas-Whitney and Grabe do not reflect the form-function-cognitive reality of Brazilian Portuguese. For example, repetitions, nominal content, third person pronouns were some of the factors loaded for the concrete-informational dimension. Though those features may relate to that function in English, they do not in BP. This dissertation demonstrates that BP has a lower type-token ratio than English because of factors that do not relate to the concrete-informational dimension. Rather, this dissertation indicates that the BP type-token ratio and related phenomena are a function of morphology and grammaticization processes. As such, they relate to cost of processing. A replication of the methodology of this study with editorials would be insightful to confirm across genre generalizations.

Attention units as measures of production and comprehension. Confirming Chafe (1994; 1988) and Chafe and Danielewicz (1987), stretches of written language between punctuation marks do seem to relate to units of attention. BP not only used shorter units than English but its punctuation units reflect constituent and clause structure with high frequency. The division of the text in clauses and constituents by punctuation seems to be a trend that deserves further investigation in BP. In that punctuation parses for the readers, their working memory may operate at other levels of the conceptualization process.

Most of all, findings with regard to length of attention units in BP seems to indicate that writing does not free writers from the constraint that keeps down the size of spoken intonation units. Contrary to Chafe and Danielewicz (1987), results of this dissertation indicate that writing also seems to limit the production of language to what can be

focused on at one time. This research not only confirmed the nine-word length for English attention units (Chafe and Danielewicz; Chafe, 1988; Chafe, 1994), but offered a potential “normal” length for attention units in BP given their frequency of occurrence. The more frequent a feature the less marked and easier to process it is (see the review on discourse processing in Chapter IV). BP attention units stayed at the length of seven words.

Apparently, as Chafe (1994) has hypothesized, length of attention units relates to the amount of information different languages pack in a word. Like Seneca’s intonation units, BP’s attention units stay at a length of seven words because one token generally encodes different conceptual information. Like the relation between amount of information in intonation units and its listeners, the amount of information in attention units seem to impact readers’ ability to shift the focus of attention. BP texts respected the “one new idea constraint” with greater frequency than their counterparts. The consistency of findings across language modalities and languages that are morphologic similar corroborates this conclusion.

It remains to be seen if spoken Brazilian Portuguese will divide intonation units in a similar way. Damron’s (1997) research on Urdu and Pakistani English departed from Chafe (1994), indicating quite the opposite. Urdu, like Brazilian Portuguese, is an inflected, more free order language that packs information within a single token. Nevertheless, its intonation units were multi-clausal, violating the “one-new idea” constraint proposed by Chafe.

The way it is, the quantitative analysis and the qualitative observation of BP texts in this corpus seem to indicate a tendency for attention units in BP to stick to the one-new idea constraint more than English. Furthermore, despite inconsistencies in punctuation,

the correlation between stretches of words bound by punctuation in BP and units of attention as measured by constituent structure or clauses occurred with high frequency. Therefore, this study offers evidence of attention units as units of production and comprehension.

Attributes of the institutional expository text image-schema override differences between speech and writing. This study ratifies Chafe and Danielewicz (1987), Prince (1981); Tannen (1982), Nystrand (1987), and Biber (1988) in that linguistic devices relate to GOAL, ORIENTATION, ALIGNMENT, and other pragmatic matters such as context, not to language modality. Some linguistic features common of this genre (institutional expository discourse) in BP and English override differences traditionally attributed to speech and writing. For example, academic papers, a genre supposedly closer to the one I analyzed, shows a much higher number of nominalizations (n=92:1000) in Chafe and Danielewicz's (1987) study. Actually, numbers of this study (n=61:1000 for English; n=65:1000 for Portuguese) are closer to those of letters and lectures in that study. Also, the number of time and locative adverbs found by this dissertation (n=37:1000) is higher than that found for academic papers (n=14:1000) and closer to the one found for letters (n=41:1000). Perhaps, because institutional expository discourse aims at selling a positive image of the companies that publish them, it exhibits features that have been typically associated with spoken discourse. Another possible reason for similarities between written institutional expository discourse and spoken language stems from the companies' need to negotiate roles with a broad readership. By moving closer to spoken language in the written-spoken continuum, addressers ease processing effort and efficiently attain the discourse goal. Like spoken language, written language is used by writers in real contexts to achieve specific goals.

Moving beyond a genre effect, BP institutional expository discourse proved to be more spoken-like than English in terms of fragmentation, length of attention units, and number of locative and time adverbials. Actually, BP institutional expository texts use attention units in a way that is closer to Chafe and Danielewicz's lectures in English, a spoken-form. BP attention units are slightly shorter (median=7 words) than the English attention units in Chafe and Danielewicz's lectures in English (mean=7.3 words). If we consider the mean as a measure of central tendency, BP attention units coincide with those of spoken English in that study (M=6).

In conclusion, there also seems to be a language effect overriding differences between spoken and written modalities. The differences between Chafe and Danielewicz' findings for English and my findings for BP discourse seem related to the way languages organize and use linguistic attributes. Their "ways with words" (Heath, 1983) go beyond genre idiosyncrasies. Rather, they aim at optimal relevance.

Research on genre. Results of this study fully confirm research on genre and sub-genres by Biber (1988; 1995) and Grabe (1987) for English, and reveal that their findings hold true for BP. Both BP and English institutional expository discourse, a sub-genre similar to annual reports and reportages in those studies, emerged as typically high informational texts. Among other things, this sub-genre typically shows a high frequency of nouns and prepositions and a high type-token ratio.

Though Biber (1988) and Dantas-Whitney and Grabe (1989) associated informational-orientation with lack of concern with the reader's reality or non-involvement, that may not be the case. According to evidence from this dissertation, despite the high informational-orientation of the texts in the corpora, there was negotiation of role-relationships with the audience (more explicit in English and more

implicated in BP). Furthermore, implicatures brought to bear by the situational context or model were also frequent and cued interaction. Therefore, the corpora analyzed emerged as informational, but also presented traces of interaction given the GOAL and ORIENTATION of discourse. This characteristic in particular may be specific of the sub-genre institutional discourse.

Other features of this sub-genre that coincide with Biber's research on register variation are frequent adverbial subordination, agent deletions, emphasis on activities, and nominalizations, and infrequent occurrence of place and time adverbials. Again, in the light of this dissertation features such as agent deletions and emphasis on activities (verbs first) in BP relate to the presence of morphology, a more free word order, and to efficient processing. In other words, it may not relate to genre but to the nature of languages and their natural logic of organization. The same interpretation may be true of *precise choice of words*. This trait proved to be truer of English than of BP institutional expository prose in this research and seems to relate to optimal relevance.

In conclusion, this study replicates existing findings about expository texts and related sub-genres. In addition, expository institutional texts in BP share general characteristics of expository institutional texts in English. However, when facts related to the presence of morphology and cost of processing come into play, they surpass genre boundaries. Results of this study offers evidence that optimal relevance comes first.

Linguistic relativity. This study endorses Hunt and Agnoli (1991) and their proposal that the Whorfian Hypothesis is about language performance rather than a cognitive and linguistic hypothesis about competence. According to findings of this dissertation, what is relative across languages is the way people combine linguistic attributes in the service of optimal, efficient communication. This is why we may find

implicated in BP). Furthermore, implicatures brought to bear by the situational context or model were also frequent and cued interaction. Therefore, the corpora analyzed emerged as informational, but also presented traces of interaction given the GOAL and ORIENTATION of discourse. This characteristic in particular may be specific of the sub-genre institutional discourse.

Other features of this sub-genre that coincide with Biber's research on register variation are frequent adverbial subordination, agent deletions, emphasis on activities, and nominalizations, and infrequent occurrence of place and time adverbials. Again, in the light of this dissertation features such as agent deletions and emphasis on activities (verbs first) in BP relate to the presence of morphology, a more free word order, and to efficient processing. In other words, it may not relate to genre but to the nature of languages and their natural logic of organization. The same interpretation may be true of *precise choice of words*. This trait proved to be truer of English than of BP institutional expository prose in this research and seems to relate to optimal relevance.

In conclusion, this study replicates existing findings about expository texts and related sub-genres. In addition, expository institutional texts in BP share general characteristics of expository institutional texts in English. However, when facts related to the presence of morphology and cost of processing come into play, they surpass genre boundaries. Results of this study offers evidence that optimal relevance comes first.

Linguistic relativity. This study endorses Hunt and Agnoli (1991) and their proposal that the Whorfian Hypothesis is about language performance rather than a cognitive and linguistic hypothesis about competence. According to findings of this dissertation, what is relative across languages is the way people combine linguistic attributes in the service of optimal, efficient communication. This is why we may find

implicated in BP). Furthermore, implicatures brought to bear by the situational context or model were also frequent and cued interaction. Therefore, the corpora analyzed emerged as informational, but also presented traces of interaction given the GOAL and ORIENTATION of discourse. This characteristic in particular may be specific of the sub-genre institutional discourse.

Other features of this sub-genre that coincide with Biber's research on register variation are frequent adverbial subordination, agent deletions, emphasis on activities, and nominalizations, and infrequent occurrence of place and time adverbials. Again, in the light of this dissertation features such as agent deletions and emphasis on activities (verbs first) in BP relate to the presence of morphology, a more free word order, and to efficient processing. In other words, it may not relate to genre but to the nature of languages and their natural logic of organization. The same interpretation may be true of *precise choice of words*. This trait proved to be truer of English than of BP institutional expository prose in this research and seems to relate to optimal relevance.

In conclusion, this study replicates existing findings about expository texts and related sub-genres. In addition, expository institutional texts in BP share general characteristics of expository institutional texts in English. However, when facts related to the presence of morphology and cost of processing come into play, they surpass genre boundaries. Results of this study offers evidence that optimal relevance comes first.

Linguistic relativity. This study endorses Hunt and Agnoli (1991) and their proposal that the Whorfian Hypothesis is about language performance rather than a cognitive and linguistic hypothesis about competence. According to findings of this dissertation, what is relative across languages is the way people combine linguistic attributes in the service of optimal, efficient communication. This is why we may find

implicated in BP). Furthermore, implicatures brought to bear by the situational context or model were also frequent and cued interaction. Therefore, the corpora analyzed emerged as informational, but also presented traces of interaction given the GOAL and ORIENTATION of discourse. This characteristic in particular may be specific of the sub-genre institutional discourse.

Other features of this sub-genre that coincide with Biber's research on register variation are frequent adverbial subordination, agent deletions, emphasis on activities, and nominalizations, and infrequent occurrence of place and time adverbials. Again, in the light of this dissertation features such as agent deletions and emphasis on activities (verbs first) in BP relate to the presence of morphology, a more free word order, and to efficient processing. In other words, it may not relate to genre but to the nature of languages and their natural logic of organization. The same interpretation may be true of *precise choice of words*. This trait proved to be truer of English than of BP institutional expository prose in this research and seems to relate to optimal relevance.

In conclusion, this study replicates existing findings about expository texts and related sub-genres. In addition, expository institutional texts in BP share general characteristics of expository institutional texts in English. However, when facts related to the presence of morphology and cost of processing come into play, they surpass genre boundaries. Results of this study offers evidence that optimal relevance comes first.

Linguistic relativity. This study endorses Hunt and Agnoli (1991) and their proposal that the Whorfian Hypothesis is about language performance rather than a cognitive and linguistic hypothesis about competence. According to findings of this dissertation, what is relative across languages is the way people combine linguistic attributes in the service of optimal, efficient communication. This is why we may find

English, on the other hand, seems to rely on strong processing instructions cued by semantic memory. This leads readers to adopt a late-closure processing strategy, mainly in the case of head-modifier sequences. Using the tools of cognitive linguistics, this study has demonstrated that head-modifier sequences in English demand re-arrangement of concepts before the final construal of the composite structure. If this is the case, English has to compensate for the late-closure strategy aiming at optimal relevance. The way English does it is by balancing the distribution of different attributes across discourse, including frequently using short sentences.

In brief, results of this study provide evidence for a psycho-pragmatic reading of the Whorfian Hypothesis. Because of morphology and word order, BP seems to favor on-line processing of discourse. English, on the other hand, because of its few morphologic resources and word order effects, seems to favor a late-closure processing strategy. It follows that each language uses linguistic attributes so that they do not bring about unjustified processing effort.

Results of the study as linguistic, social, and cultural phenomena. On one hand, this study offers some evidence of socio-cultural influences in the way BP organizes expository discourse. As I mentioned before, the presence of long sentences, frequent embedding, and right-branching constructions has been interpreted in contrastive analysis of discourse (e.g. Oliveira, 1997a; 1997b; Lux and Grabe, 1991; Reppen and Grabe, 1993) as marks of elaboration grounded in literacy traditions and cultural values.

In the case of BP, this elaborated style is indeed the style of public communication and it is the style Brazilian speech communities respect, mainly the academic community. Historically, its roots go back to a European tradition of literacy that nurtures and values elaborated rhetoric and that has served as a model for the Brazilian educational system

and society at large. Thus, the results of this study in terms of social and cultural phenomena adds to findings related to Spanish (Reid, 1988; Lux, 1991; Lux and Grabe, 1991; Reppen and Grabe, 1993); French (Silva, 1992), German (Clyne, 1991); and Dutch and German (Duzack, 1995) and confirm Oliveira (1997a; 1997b). They also build on the robust corpus of evidence for the typicality of long sentences in Spanish (Reid, Lux, Lux and Grabe, Reppen and Grabe) and Italian (see Bates and McWhinney, 1989), languages that belong to the Romance family and that share structural similarities with BP.

On the other hand, it may also be that the so-called “elaborated” style of Portuguese, as I discussed before, stem from what grammar allows *vis-à-vis* optimal relevance. In other words, it is because BP grammar allows users to process on-line that they can produce and comprehend lengthy sentences, frequent embedding and right branching constructions. What I have been trying to stress is that the way languages organize linguistic attributes in texts appears to reflect how and what grammar allows them to process. This interpretation gains strength if we consider that the structure of the languages that founded those very literacy traditions is similar to Brazilian Portuguese. One of them is French. It gains even more strength if we consider that languages that share different literacy traditions such as German and Dutch offer similar evidence. In my view, grammar seems to play a considerable role. The forum, though, is still open. It may be that both grammar and cultural values are at play. Future research should investigate the issue.

Translation studies. Translation proved to be a tool for discourse analysis. In conducting the qualitative analysis, some pragmatic and cognitive details only came to light through the analysis of translated texts. The comparison between the BP source

texts and their English variants strengthened the understanding that variants of a text form a family of resemblances within a image-schema. The variants fall at different points in the COMMUNICATIVE TEXT image-schema depending on the attributes that co-occur and on their frequency of occurrence. The more variants reveal attributes of the COMMUNICATIVE TEXT the closer they are to the prototypical exemplar. If few attributes co-occur, we have a marginal exemplar and a translation that hardly conveys the image-schema of the source text.

Finally, the examination of BP texts and their English translation variants also confirmed Toury (1991), Gentzler (1993), Tirkkonen-Condit (1989), Kemppinen (1988), and Weise (1988). There are different degrees of adequacy and equivalence. The more texts in the target language share attributes of the COMMUNICATIVE TEXT in that language, the more adequate and equivalent they are to texts in the source language. Typicality of linguistic attributes combines with rhetoric and pragmatic factors to construe texts that belong to the target language family of texts. The more those factors are taken in consideration in the translation process, the more equivalent and adequate the translation will be.

Furthermore, this investigation adds that typicality of linguistic attributes should also combine with cognitive and perceptual factors to guide translators in their tasks. The type of processing instruction cued by grammar in distinct languages is one important consideration to bear in mind. Sentence length is another. Sentences in the source texts were systematically broken in two, even three sentences in the target texts. In doing so, the translator allowed autonomous entities to appear in a perceptually salient position, drew on the few morphologic resources of English, and compensated for word-order effects on the allocation of attention resources. Lexical organization and the number of

independent constructions strengthen this contribution as well. Translation used an even more varied range of vocabulary in collocation networks than English as well as a higher number of independent constructions. The slow but strong processing instructions typical of English image-schema were respected.

If translators fail to consider how readers allocate attention resources in the target language, they run the risk of construing variants that will hardly be understood or fully appreciated. This seems to have been the case of Dittmar's *Sociolinguistik* (cited in Clyne, 1991) that got such a negative review in America because of its pretentious style and bad organization. Actually, the translation to English failed to transform the "elaborated" German style into English "reduced" style. This also seems to have been the case of texts in this study that overused head-modifier sequences or that used attention units, juxtaposed PPs, and non-finite clauses in a BP-like manner. These cases led us to conclude that translation discourse in this study emerged as a fuzzy category, appearing at the boundaries of English and BP image-schemas. The way in which these cases used the attributes probably impaired the allocation of attention resources by many readers.

Finally, the review of translation studies, my own analysis of English translation texts in the corpus, and the cognitive paradigm strongly emphasize that translation is a "matter of different modes of cognitive processing" (Lakoff, 1990, p.320) and that the *locus of variation* (p. 329) is in the linguistic system itself. While English employs a late-closure strategy and relies on semantic memory to construe a message, Portuguese relies on on-line processing. Those distinct processing strategies arise from the language specific attributes present in texts in each language. To map BP texts onto the domain of

the English COMMUNICATIVE TEXT, the translator of the corpus used and organized attributes in ways that reflected that change in processing modes.

General comments. Evidence provided by this study assigns prominence to bottom up processes. However, the role of top-down processes is recognized in that it facilitates inferences and enlarges the capacity of working memory. In the light of that understanding, Gernsbacher's model of discourse processing (1996)—the Structure Building Framework model—seems to fit this data adequately. Also, though the study did not address the affective variables related to text production or the writing process itself, Hayes (1996) and Grabe and Kaplan (1996) offer socio-cognitive models that seem to fit the evidence provided by this data. Both models recognize the limitations of working memory in the production of writing. Other components of these two models as well as differences in the way they fit the data cannot be brought into discussion given that I did not investigate the writing process but its results.

CHAPTER VIII

CONCLUSION

This dissertation combined the descriptive tools of discourse analysis with those of Cognitive Grammar as conceived by Ronald Langacker (1987; 1991) to contrast Brazilian Portuguese institutional expository prose to English, and to provide a cognitive rationale for discourse translation. It uncovered attributes that are typical of that genre in English and Brazilian Portuguese and proposed an image-schema for the COMMUNICATIVE TEXT category in each language. It has also demonstrated that those schemas emerge from the users' need to balance cost of processing aiming at optimal relevance. Finally, it explained discourse translation as a change in mode of processing brought to bear by attributes of the COMMUNICATIVE TEXT in different languages. The analysis linked linguistic competence with performance by calling on the limitations of the human attention resources and other psychological mechanisms that may also explain rhetorical differences.

Theoretical conclusions. That the devices different languages use in written discourse conjoin to optimize cost of processing is, as far as I can tell from the literature available, a relatively new idea, and one that will prove valuable in the study of contrastive analysis of discourse and of the role the mind plays in shaping written language. This leads me to a second contribution of this study of expository discourse in Portuguese and English. The ways the two languages use the selected linguistic events corroborate Hunt

and Agnoli's (1991) psychological explanation of the Whorfian Hypothesis. They use and organize the attributes in such a way as to counterbalance different modes of processing favored by their grammars. This contribution is a significant step toward a better understanding of an issue that has been in the linguistic forum for decades. Cost of processing or optimal relevance, in Sperber and Wilson's (1986, 1995) terminology, seems to be at the root of contrastive rhetoric patterns.

Finally, human language communication occurs with the purpose of being understood and understanding is a cognitive process (Johnson, 1987). The COMMUNICATIVE TEXT is strong evidence that an objectivist-rationalist view of cognition does not hold. There is no way all readers will find one and only one optimally relevant interpretation or correspondence between the linguistic and mental representations because readers come from different environments and the process of category formation is individual. Therefore, only a subjective-empirical view of cognition provides explanatory adequacy for discourse translation and for variation within the COMMUNICATIVE TEXT image-schema.

According to the model I here propose, a number of cognitive categories combine to form the COMMUNICATIVE TEXT image-schema. Members diverge in the frequency of use and weight of one or more attributes in the model. However, the overall structure resembles that of the image-schema as brothers and sisters resemble other relatives in the family. Though I have worked briefly with variation within categories, the distribution of attributes across different texts in the sample attests the variation. It illustrates not only the more central member of English and BP institutional expository discourses, but also the spread of the cognitive categories.

Methodological conclusions. The study has been interdisciplinary with greater focus on cognitive linguistics, discourse analysis, pragmatics, contrastive rhetoric, and cognitive psychology, in that order. The notion of texts as schemas and the category COMMUNICATIVE TEXT have nourished from developments in those disciplines and has proved successful and insightful in that it allowed me to offer a cognitive-pragmatic model for contrastive analysis of discourse.

However, I need to caution readers to be conservative about results related to BP. Textual analysis is only an indirect tool to investigate psychological processes. Since research in the field has neglected on-line experiments on sentence processing in Portuguese, future research should try to observe and investigate the phenomenon in natural settings, as much as plot the number of interpretations as a function of sentence and intonation unit length. In other words, to further validate this analysis, future studies should focus on the cognitive strategies at work during BP discourse production.

Furthermore, because institutional texts fulfill a communicative function—to sell a positive image of the companies they advertise and that publish them—you may be asking yourself about the generalizability of my results to other sub-types of expository texts or even other genres. My answer to this question will follow two venues: one has to do with the COMMUNICATIVE TEXT image-schema, the other with the results of the empiric study. Nothing precludes the COMMUNICATIVE TEXT from being applied to the description of different genres. The model is generic and arose from my study of different theoretical notions and studies that exist in discourse analysis, cognitive psychology, psycholinguistics, and cognitive linguistics. The type of linguistic description the model

motivates undoubtedly provides a more accurate view of texts as conceptualized by people in their cognitive, socio-cultural universes.

Besides, I strongly suspect that findings of the empirical study such as sentence and attention unit length, use of independent versus non-finite constructions, and lexical organization will apply to research of this kind across genres in the two languages. The typical use of lengthy sentences in BP and short sentences in English has already been confirmed by research on expository prose (see chapter II). Short sentences in English have also appeared in narratives.

Finally, my last and final argument for generalizability across genres rests on the effects of word order and the relation between grammar and processing instructions. The cognitive effects of head-modifier relations are facts that derive from the structure of English and Brazilian Portuguese and that certainly hold across genres. Cognitive factors, such as on-line or a late-closure processing, stem from anaphoric ties, mainly verb and noun inflections; from the organization of the lexicon, mainly collocation of content words; and from the perceptual salience of autonomous entities. These are also structural-cognitive facts about the two languages that will equally hold across genres.

Thus, I suggest that the COMMUNICATIVE TEXT image-schema, as conceived in Chapter IV of this dissertation (Figures 5 and 6), is not genre specific, but a general model of linguistic description. I also suggest that the cognitive and perceptual factors of the PATH sub-component mentioned above hold across genres. Tendencies established for the frequency of selected attributes will probably also hold. Frequency of use derives from cognitive factors. As for the other attributes the empirical study address, future research on other genres may strengthen my findings and provide a definite answer to any doubt

about generalizability of results. Future studies should also investigate other attributes of the schema in each language.

Another question that you as a reader may have is if the English used in these brochures is not different from U.S. English, given that the English brochures aim at an international audience. If the English is different, the BP also is. The BP brochures are also read by Spanish speakers and English speakers who have lived in Brazil, and, who speak Portuguese as a second language. Thus, they also aim at an international audience. The brochures primary audience, however, are native speakers. Otherwise, they would not have variants in so many other languages. These variants are available from the oil companies.

Implications

Texts as a COMMUNICATIVE category implies concern with aspects of human communication that can only be captured through a situated, cognitive linguistic analysis of language in its stream of life and from different perspectives. Written information drives the globalized world we live in and promote intercultural communication. The more we are able to understand the processes that govern its efficiency, the more successful we will be in our ventures.

Though, as Enkvist (1987) says, we are still far from the point where we may relate successes and failures in composition directly to text processing, texts look the way they do because they have to be processed in real time. Therefore, this study bears direct applications to composition studies and writing pedagogy in first and foreign languages and to traditional treatments of error in language acquisition. Particularly, it emphasizes that writing is also subject to the limitations of language production. If writers were

aware of those limitations, they could better monitor the writing process, and manage their own attention resources. An efficient management of working attention at the time of production will certainly originate an efficient comprehension process and attain the GOAL of written discourse: communication.

The analysis of the cognitive sub-components of events such as sentence length, lexical collocation or word order may help non-natives to understand the motivation behind the English more SVO word-order or behind any other attribute that sharply contrasts with the native language and potentially causes an array of barriers in the writing classroom. This may help native speakers as well. Teaching students about how linguistic attributes combine in more complex gestalts to accomplish communicative goals and why those attributes, if combined appropriately, may help them in their tasks should improve the efficiency of the writing classroom, the writing process, and the writing products.

Furthermore, the treatment of grammatical phenomena, especially discourse, within the cognitive linguistics paradigm, provides a more plausible account of errors in foreign language acquisition. It is when dealing with marginal members of the categories that students get in trouble. Normally, those members that are at the borderline have so few connections with the overarching image-schema that students get lost and wander through other domains. Readers, for their part, can hardly retrieve the intended message. If language users are instructed about central instances of different categories in the target language, they will be better equipped to construe more prototypical members of those categories.

In a similar vein, this study also bears direct applications to translation studies. The more cross-linguistic studies investigate typical attributes across genres or language

specific text image-schemas the more translators will know how to use and organize linguistic attributes to yield maximal cognitive effects. If translators adjust their choices to conform to choices natives would rapidly recognize as attributes of the COMMUNICATIVE TEXT, they will produce texts that natives process with less effort and greater speed. In other words, translators will improve the acceptance of their work.

Finally, the implications of my analysis for contrastive discourse analysis strongly suggest a reconsideration of existing accounts. According to this study, languages have different frequencies of linguistic events because events instruct addressers about conceptualization routes. Given that each language has specific structural characteristics, they necessarily yield different frequencies of attributes when in textual combination as well as different rhetorical patterns. Again, cognitive effects are the roots of different patterns.

Summary of contributions

This study breaks through in that it

1. Adopts an interdisciplinary thrust. The descriptive tools of CG were combined with principles of cognitive linguistics, cognitive psychology, and the tools of discourse analysis to describe discourse and link linguistic competence to performance. There are no other studies that have ever done that.
2. Proposes an image-schema: the COMMUNICATIVE TEXT (Figures 5 and 6, Chapter IV).
3. Uncovers typical attributes of English and BP institutional expository prose.
4. Proposes an image-schema for the COMMUNICATIVE TEXT in each language as well as explains variation within the categories (Figures 19 and 20, Chapter VII).

5. Provides a cognitive explanation for the distinct organization of similar attributes into discourse across languages. Linguistic attributes conjoin to balance cost of processing.
6. Provides a cognitive explanation for discourse translation.
7. Proposes a psycho-pragmatic model for contrastive analysis of discourse.

Recommendation for further study

In the last 15 years, much has been learned about the frequency of linguistic attributes in expository and narrative discourses in English and a few other languages. Much has also been learned about discourse processing in languages such as English, French, Italian, Spanish, German, and Chinese. However, there is still a lot to be accomplished in the fields of discourse analysis, contrastive rhetoric, and discourse processing. To begin with, it appears to me that research in the field of discourse analysis would benefit enormously if researchers would use the descriptive tools of Cognitive Grammar. Cognitive Grammar would clarify information on certain notions such as cognitive routes and markedness theory. Second, my review of the literature seems to indicate that research across genres in different languages has been neglected. Most information available relates to English. Biber (1995), Dantas-Whitney and Grabe (1989), and Oliveira (1997a;1997b) are some of the few contributions that address specific genres in different languages. Actually, published studies about Portuguese discourse are nearly non-existing (Lux and Grabe, 1991). Most information comes from unpublished dissertations and studies that have been written in Brazilian Portuguese and that are thus of limited access for the international audience. Further investigations of Brazilian Portuguese would facilitate the observation of typical cues in distinct genres (or sub-

genres) and ultimately the work of writers, translators, and intercultural communication. Furthermore, it would also enlarge our knowledge of how languages work in discourse. Third, future studies could further elucidate facts about BP, its attributes, and their relation to effort of processing. One interesting aspect to investigate is if BP intonation units will stick to the one-new idea constraint as Chafe (1994) has proposed for spoken language. As I mentioned previously, only on-line studies can validate the analysis and conclusions I offer here. Another aspect of BP for future investigation has to do with clause position and the foreground-background asymmetry. While conducting the qualitative analysis for this study, I have noticed that some subordinate clauses in BP seem to foreground information. Fourth, investigation of attributes such as passive constructions and manner adverbs; differences and similarities between the use and frequency of non-finite clauses and juxtaposed PPs in BP may also enrich the institutional “communicative BP text” category here proposed. Fifth, future research on BP should investigate topic development. I have also noticed that verbs tend to appear first with great frequency and give rise to new sub-topics. In other words, apparently BP tends to use different themes that amplify a hyper-theme previously established. Verbs as prominent structures in discourse may add another cognitive/perceptual dimension to the ways languages organize and use attributes. Why ACTION first? Most of all, future studies could examine the role of punctuation in different languages and corroborate its relation to production and comprehension. One aspect research of this kind should address is if attention units across languages hold to the one-idea constrain and how they compare to findings in spoken language.

Conclusion

The biggest quality of this type of analysis is that it gives us a theory of contrastive analysis of discourse that links competence with performance, form with semantic and pragmatic functions, production with cognition. Written discourse becomes the grammatical representation of an image-schema as perceived by the writer and stored in the cognitive grammar of language users. That is, interactions of form and function simply replicate the allocation of attention resources in discourse comprehension (Tomlin, 1995) and production. Only through the study of discourse across various languages can we sort the role of various linguistic devices in cognition, and more precisely, in text processing.

REFERENCES

- Aaronson, D., and Ferres, S. (1986). Sentence processing in Chinese-American bilinguals. *Journal of Memory and Language*, 25, 136-162.
- Abu-Ssaydeh, A. F. (1993). Business translation: A personal perspective. *Babel*, 39, 1-10.
- Akinaso, F. N. (1982). On the differences between spoken and written language. *Language and speech*, 25 (1), 97-125.
- Akmajian, A., Demers, R.A., Farmer and Harnish, R.M. (1990). *An introduction to language and communication* (3rd. edition). Cambridge, MA: MIT Press.
- Almeida, J.C.P. (1984). The interplay of cohesion and coherence in native and non-native written academic discourse. *Dissertation Abstracts International*, 4 (5), 1263^A (University Microfilms No. 86-15, 681)
- Alvarez, A. (1993). On translating metaphor. *Meta*, 38, 479-490.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409-429.
- Anderson, J.R. (1995). *Cognitive psychology and its implications* (4th. edition). New York: W.H. Freeman and Company.

- Anderson, A., Garrod S.C., and Sanford, A.J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, 35A, 427-440.
- Armstrong, S.L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13, 263-308.
- Atari, O. (1984). *Oral style strategies in EFL written discourse: Implications for teaching college composition*. Paper presented at the 3rd. Annual Linguistics Conference in Jordan. (ERIC Document Reproduction Service No. ED 271 006).
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (1990). *Human memory: Theory and practice*. Boston: Allyn and Bacon.
- Bahrick, H.P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1-24.
- Barbe, Katherina. (1992). *Translation and Text Analysis*. Paper presented at the 1992 Annual International Conference on Pragmatics and Language Learning. (ERIC Document Reproduction Service No. ED 349 554).
- Barsalou, L. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: L. Erlbaum.
- Basso, K.H. (1974). The ethnography of writing. In R. Bauman and J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 425-32). Cambridge: Cambridge University Press.

- Bates, E., and MacWhinney, B. (1982). Functionalist approaches to grammar. In Wanner, E., & Gleitman, L. (Eds.), *Language acquisition: the state of the art*. New York: Cambridge University Press.
- Bates, E., and MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3-73). Cambridge: Cambridge University Press.
- Bates, E., MacWhinney, E., Caselli, C., Devescovi, A., Natalie F., and Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child development, 55*, 341-354.
- Bates, E., McNew, S., MacWhinney, B., Devescovi, A., and Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition, 11*, 245-299.
- Beaugrande, R. de. (1980). *Text, discourse, and process: Toward a multi-disciplinary science of texts*. Norwood, NJ: Ablex. [Advances in Discourse Processes, Vol. IV.]
- Beaugrande, R. de, and Dressler, W. (1981). *Introduction to text linguistics*. London: Longman.
- Beaugrande, R. de. (1990). Text linguistics through the years. *Text 10* (1/2), 9-17.
- Becker, A. L. (1980). Text-building epistemology, and aesthetics in Javanese Shadow Theatre. In A. Becker and A. Yengoyan (Eds.), *The imagination of reality*. Norwood, N.J.: Ablex.
- Biber, Douglas. (1984). A model of textual relations within the written and spoken modes. Unpublished doctoral dissertation, University of Southern California.
- Biber, Douglas. (1986). Spoken and written textual dimensions in English: resolving the contradictory findings. *Language, 62*, 384-414.

- Biber, Douglas. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Blackmore, Diane. (1992). *Understanding utterances: An introduction to pragmatics*. Cambridge: Blackwell.
- Blass, Regina. (1990). *Relevance Relations in Discourse: A Study with special reference to Sissala*. Cambridge: Cambridge University Press.
- Bock, J.K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1-47.
- Bornetto, C. Serra. (1996). *Liegen and stehen in German: A study in horizontality and verticality*. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.459-506). New York: Mouton de Gruyter.
- Bower, G.H., Clark, M., Winzenz, D., & Lesgold, A. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 8, 323-343.
- Bransford, J.D., and Johnson, M.K.. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal learning and Verbal Behavior* 11, 717-726.
- Bransford, J. D. and Johnson, M.K. (1973). Considerations of some problems of comprehension. In W.G. Chase (Ed.), *Visual information processing*. Orlando, FL: Academic Press.

- Brennan, Mark, and Brennan, Roslin. (1990). *Strange language: Child victims under cross examination*. Australian Institute of Criminology. Criminology Research Council (3rd. edition; chapter IX, pp. 60-66). ISBN 0 949060-74-7. (ERIC Document Reproduction Service ED 336 950).
- Brewer, W.F., and Treyens, J.C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207-230.
- Brewer, W.F., and Nakamura, G.V. (1984). The nature and function of schemas. In R.S. Wyer and T.K. Srull (Eds.), *Handbook of social cognition*. Hillsdale, NJ: Erlbaum.
- Britton, B. (1994). Understanding expository text. In M.A.Gernsbacher (Ed.) *Handbook of psycholinguistics* (pp.641-674). New York: Academic Press.
- Britton, B., van Dusen, L., Glynn, S.M., and Hemphill, D. (1990). The impact of inferences on instructional text. In A.C.Graesser and G.H. Bower (Eds.), *Inferences and text comprehension*. San Diego: Academic Press.
- Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.
- Brown, G., and Yule G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Bybee, Joan, and Moder Carol. (1983). Morphological classes as natural categories. *Language*, 59 (2), 251-70.
- Camara, Mattoso Jr. (1977). *Estrutura da língua portuguesa* [The structure of Portuguese]. Petrópolis, RJ., Brasil: Editora Vozes.
- Caille, Pierre-François. (1974). Translators and translation: 1974 Survey. *Babel*, 20 (3), 130-141.

- Carpenter, P., Miyake, A., and Just, M. (1994). Working memory constraints in comprehension. In M.A.Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.1075-1122). New York: Academic Press.
- Carrell, P.L. (1982). Cohesion is not coherence. *TESOL Quarterly*, 16 (4), 479-488.
- Carrell, P.L. (1983). Some issues in studying the role of schemata, or background knowledge in second language comprehension. *Reading in a foreign language* 1 (2), 81-92.
- Carrell, P.L. (1987). Text as interaction: Some implications of text analysis and reading research for ESL composition. In Ulla Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*, 57-71. Reading, MA: Addison-Wesley.
- Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, 42, 1-20.
- Casad, Eugene. (1988). Conventionalization of Cora locationals. In Brygida Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (345-378). Amsterdam and Philadelphia: John Benjamins.
- Casad, Eugene. (1995). Seeing it in more than one way. In J.R. Taylor and R.Mac Laury (Eds.) *Language and the cognitive construal of the world* (pp.23-50). New York: Mouton de Gruyter.
- Casad, Eugene. (1996). Introduction. In E.H.Casad (Ed.) *Cognitive linguistics in the redwoods* (pp.6-26). New York: Mouton de Gruyter.
- Catford, J. C. (1974). *A linguistic theory of translation*. London: Oxford.
- Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review* 79, 525-530.

- Celce-Murcia, Marianne. (1990). Discourse analysis and grammar instruction. *Annual Review of Applied Linguistics*, 11, 135-151.
- Chafe, Wallace L. (1973). Language and memory. *Language*, 49, 261-281.
- Chafe, Wallace L. (1974). Language and consciousness. *Language*, 50, 11-33.
- Chafe, Wallace L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N.Li (Ed.), *Subject and topic*. New York: Academic Press.
- Chafe, Wallace L. (1982). Integration and involvement in speaking, writing, and oral literature. In Deborah Tannen (Ed.), *Spoken and written language: Exploring orality and literacy*. Norwood, N.J.: Ablex.
- Chafe, Wallace L. (1985). Linguistic differences produced by differences between speaking and writing. In D.R.Olson, N.Torrance, and A.Hildyard (Eds.), *Literature, language, and learning: the nature and consequences of reading and writing* (pp. 105-23). Cambridge: Cambridge University Press.
- Chafe, Wallace L. (1987). Cognitive constraints on information flow. In Russell Tomlin (Ed.), *Coherence and grounding in discourse* (pp.21-51). Amsterdam and Philadelphia: John Benjamins.
- Chafe, Wallace L. (1988). Punctuation and the prosody of written language. *Written Communication*, 5, 396-426.
- Chafe, Wallace L. (1990). Some things that narratives tell us about the mind. In Bruce K. Britton and Anthony Pellegrini (Eds.), *Narrative thought and narrative language*. Hillsdale, N.J.: Lawrence Erlbaum.

- Chafe, Wallace L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.
- Chafe, W., & Danielwicz, J. (1987). Properties of spoken and written language. In R. Horowitz & F.J. Samuels (Eds.), *Comprehending oral and written language* (pp.83-112). New York: Academic Press.
- Charolle, M. (1983). Coherence as a principle in the interpretation of discourse. *Text*, 3 (1), 71-97.
- Charrow, R.P. and Charrow, V.R. (1979). Making legal language understandable: a psycholinguistic study of jury instructions. *Columbia Law Review*, 79, 1306-1374.
- Charvillat, A. and Kail M. (1991). The status of 'canonical SVO sentences' in French: a developmental study of the on-line processing of dislocated sentences. *Journal of Child Language* 18, 591-608.
- Clark, H. H., and Clark, E.V. (1977). *Psychology and Language*. New York: Hartcourt, Brace, Jovanovich.
- Clark, H.H., and Haviland, S.E. (1977). Comprehension and the Given-New contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1-40). Norwood: Ablex.
- Clyne, M. (1981). Culture and discourse structure. *Journal of Pragmatics*, 5, 61-66.
- Clyne, M. (1987). Discourse structures and discourse expectations: Implications for Anglo-German academic communication in English. In Larry Smith (Ed.) *Discourse across cultures* (pp. 73-100). Prentice Hall.

- Clyne, M. (1991). The sociocultural dimension: The dilemma of the German-speaker scholar. In H. Schroder (Ed.) *Subject-oriented texts* (pp. 49-67). New York: Walter de Gruyter.
- Coleman, L., and Kay P. (1981). Prototype semantics: the English word *lie*. *Language* 57, 26-44.
- Connor, U. (1984). A study of cohesion and coherence in ESL students' writing. *Papers in Linguistics: International Journal of Human Communication*, 17 (3), 301-316.
- Connor, U. (1987). Argumentative patterns in student essays: Cross-cultural differences. In Ulla Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*, 57-71. Reading, MA: Addison-Wesley.
- Connor, U. (1996). *Contrastive rhetoric*. Cambridge: Cambridge University Press.
- Connor, U., & Kaplan, R (Eds.). (1989). *Writing across languages: Analysis of L2 text*. Reading, MA: Addison-Wesley.
- Coulthard, Malcom, (1994). *Advances in written text analysis*. London: Routledge.
- Cuetos, F., and Mitchell, D.C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, 30, 73-105.
- Cunha, Celso, and Cintra, L.F. (1985). *Nova gramática do português contemporâneo* [A new grammar of contemporary Portuguese] (2nd edition). Rio de Janeiro, Brasil: Editora Nova Fronteira.
- Damron, Rebecca. (1997). *Prosody in Urdu and Pakistani English conversational discourse*. Unpublished doctoral dissertation, Oklahoma State University, Stillwater, OK.

- Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F.Daneš (ed.), *Papers on Functional Sentence Perspective* (pp. 106-128). Prague: Academia.
- Danet. (1980). Language in the legal process. *Law Society Review*, 14 (3), 445-555.
- Dantas, Francisco de Assis. (1983). *A Ordem das Palavras em Português* [Word-order in Portuguese]. Unpublished Master Thesis. Pontifícia Universidade Católica do Rio de Janeiro, Brasil.
- Dantas-Whitney, M., and Grabe, W. (1989). *A comparison of Portuguese and English newspaper editorials*. Paper presented at the 23rd. Annual TESOL Convention, San Antonio, Texas.
- Davis, Steven. (1991). *Pragmatics: A reader*. New York: Oxford University Press.
- Deanne, Paul. (1996). Neurological evidence for a cognitive theory of syntax: Agrammatic aphasia and the spatialization of form hypothesis. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.55-116). New York: Mouton de Gruyter.
- Delbecque, Nicole. (1996). Towards a cognitive account of the use of the prepositions *por* and *para* in Spanish. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp. 249-318). New York: Mouton de Gruyter.
- Delia, J.G., Kline, S.L., and Burleson B.R. (1979). The development of persuasive communication strategies in kindergarteners through twelfth-graders. *Communication monographs*, 46, 241-256.
- Delisle, J. (1988). *Translation: An interpretative approach*. Ottawa, Canada: University of Ottawa Press.

- Doughty, Anne, & Doughty, Peter. (1975). *Language and Community*. London: Edward Arnold.
- Doherty, M. (1992). Relativity of sentence boundary. *Babel*, 38, 72-78.
- Duszak, Anna. (1995). On variation in news-text prototypes: Some evidence from English, Polish, and German. *Discourse Processes*, 19, 465-483.
- Eggington, William G. (1987). Written academic discourse in Korean: Implications for effective communication. In Ulla Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*, (pp. 153-69). Reading, MA: Addison-Wesley.
- Enkvist, Nils E. (1987). Text Linguistics for the applier: An orientation. In Ulla Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text*, 23-43. Reading, MA: Addison-Wesley.
- Erikson, C.W., and St. James J.D. (1986). Visual attention within and around the field of focal attention: A zoom lens models. *Perception and Psychophysics*, 40, 225-240.
- Erlich, S. (1988). Cohesive devices and discourse competence. *World Englishes*, 7, 11-118.
- Erteschik-Shir, N., and Lappin, S. (1979). Dominance and the functional explanation of island phenomena. *Theoretical Linguistics*, 6, 41-86.
- Faigley, L. (1980). Names in search of a concept: maturity, fluency, complexity, and growth in written syntax. *College Composition and Communication*, 31 (3), 291-300.
- Fauconnier, G. (1994). *Mental spaces*. New York: Cambridge University Press.
- Fillmore, C. J. (1968). Remarks on contrastive pragmatics. In J. Fisiak (Ed.), *Contrastive linguistics prospects and problems*, (pp. 119-141). New York: Mouton Publishers.

- Firbas, Jan. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.
- Fletcher, C.R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language*, 25, 43-58.
- Fletcher, C. R. (1994). Levels of representation in memory for discourse. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.589-607). New York: Academic Press.
- Flores d'Arcais, G.B. (1987). Perceptual factors and word order in event descriptions. In G. Kempen (Ed.), *Natural language generation: New results in artificial intelligence, psychology, and linguistics*, (pp. 441-451). Dordrecht: Martinus Nijhoff.
- Flores d'Arcais, G.B., and Schreuder, R. (1987). Semantic activation during object naming. *Psychological Research*, 49, 153-159.
- Floyd, Rick. (1996). The radial structure of the Wanka reportative. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.895-942). New York: Mouton de Gruyter.
- François, Denise. (1974). *Français parlé. Analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne* [Spoken French: analysis of data collected in the area of Paris]. (2 vol). Paris: S.E.L.A.F.
- Frazier, L., and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Garrod, S.C., and Sanford, A.J. (1985). On the real time character of interpretation during reading. *Language and Cognitive Processes*, 1, 43-50.

- Garrod, S.C., and Sanford, A.J. (1990). Referential processes in reading: Focusing on roles and individuals. In G.B. Flores d'Arcais, K. Rayner, and D. Balota (Eds.), *Comprehension Processes in Reading*. Hillsdale, NJ: Erlbaum.
- Gernsbacher, M. A. (1995). Activating knowledge of fictional characters' emotional states. In C. Weaver, S. Mannes, and C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp.141-155). Hillsdale, NJ: L. Erlbaum.
- Gernsbacher, M.A. (1996). The structure-building framework: What it is, what it might also be, and why. In B.Britton and A.Graesser (Eds.), *Models of understanding text* (pp. 289-311). Mahwah, NJ: L. Erlbaum.
- Gernsbacher, M.A., and Givón, T. (1995) *Coherence in spontaneous text*. Amsterdam and Philadelphia: John Benjamins.
- Gernsbacher, M.A., and Shroyer, S. (1989). The cataphoric use of the indefinite this in spoken narratives. *Memory and Cognition*, 17, 536-40.
- Gernsbacher, M.A., and Hargreaves. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27, 699-717.
- Gentzler, E. (1993). *Contemporary translation theories*. New York: Routledge.
- Gibbs, Raymond. (1996). What's cognitive about cognitive linguistics? In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.27-54). New York: Mouton de Gruyter.
- Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp.1-42). Amsterdam and Philadelphia: John Benjamins.

- Givón, T. (1987). Beyond foreground and background. In Russel Tomlin (Ed.), *Coherence and grounding in discourse* (pp.175-188). Amsterdan and Philadelphia: John Benjamins.
- Givón, T. (1989). *Mind, code, and context: Essays in pragmatics*. Hillsdale, NJ: Erlbaum.
- Givón, T. (1990). *Syntax: A functional-typological introduction* (Vol. II). Philadelphia: J. Benjamins.
- Givón, T. (1993). *English grammar: A function-based introduction* (Vol. II). Bethlehem Pike, PA: John Benjamins.
- Givón, T. (1995). Coherence in text vs. coherence in mind. In M.A.Gernsbacher and T. Givón (Eds.), *Coherence in spontaneous text* (pp.59-115). Philadelphia: J.Benjamins.
- Goldman and Varna. (1995). CAPPING the Construction-integration model and other inventions. In C. Weaver, S. Mannes, and C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp.337-358). Hillsdale, NJ: L. Erlbaum.
- Grabe, William. (1987). Contrastive rhetoric and text-type research. In Ulla Connor and R.B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 115-137). Reading, MA: Addison-Wesley.
- Grabe, William (Ed.). (1990). *Annual Review of Applied Linguistics 11*. New York: Cambridge Unviersity Press.
- Grabe, William. (1996). *Textual coherence and discourse analysis*. Paper presented at the 1996' AILA, July 1996.

- Grabe, W., & Kaplan, R. (1989). Writing in a second language: Contrastive rhetoric. In Donna Johnson and Duane Roen (Eds.), *Richness in writing: Empowering ESL students*. Longman: New York.
- Grabe, William, and Kaplan, R. (1996). *Theory and practice of writing*. New York: Longman.
- Graesser, A., and Britton, B. (1996). Five metaphors for text understanding. In B. Britton and A. Graesser (Eds.), *Models for understanding text*(341-351). Mahwah, NJ: L. Erlbaum.
- Graesser, A., and Zwaan. (1995). Inference generation and the construction of situation models. In C. Weaver, S. Mannes, and C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 117-139). Hillsdale, NJ: L. Erlbaum.
- Graf, P., Squire, L.R., and Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 10, 164-178.
- Greenbaum, S., and Quirk, R. (1991). *A student's grammar of the English language*. Essex, England: Longman Group.
- Grice, H.P. (1975). Logic and conversation. In P.Cole and J.Morgan (Eds.), *Speech acts (Syntax and semantics, Vol. 3, pp. 41-58)*. New York: Academic Press.
- Grice, H.P. (1989). *Studies in the way of words*. Cambridge, Mass.: Harvard University Press.
- Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Gumperz, J. (1984). Communicative competence revisited. In Schiffrin (Ed.) *Meaning, form and use: Linguistic applications* (278-289). Washington DC: Gerogetown

- University Press. [Georgetown University Round Table on languages and linguistics 1984].
- Györi, Garbor. (1996). Historical aspects of categorization. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.175-206). New York: Mouton de Gruyter.
- Haiman, John, and Thompson, Sandra. (1984). "Subordination" in Universal Grammar. *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society* (pp. 510-523).
- Hall, J. F. (1954). Learning as a function of word frequency. *American Journal of Psychology*, 67, 138-140.
- Halliday, M.A.K. (1967). Notes on transitivity and theme in English. Parts 1 and 2. *Journal of Linguistics*, 3, 37-81, 199-244.
- Halliday, M.A.K. (1970). *A course in spoken English: intonation*. Oxford: Oxford University Press.
- Halliday, M.A.K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, M. A.K. (1979). Differences between spoken and written language: Some implications for literacy teaching. In G. Page, J. Elkins, and B. O'Connor (Eds.), *Communication through reading: Proceedings of the 4th. Australian Reading Conference* (vol. 2., pp. 37-52). Adelaide, S.A.: Australian Reading Association.
- Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Edward Arnold.

- Halliday, M.A.K. (1987). Spoken and written modes of meaning. In Rosalind Horowitz and S. Jay Samuels (Eds.), *Comprehending oral and written language* (pp.55-81). New York: Academic Press.
- Halliday, M.A.K. (1989). *Spoken and written language*. Oxford: Oxford University Press.
- Halliday, M.A.K., and Hasan, R. (1976). *Cohesion in English*. Essex, England: Longman.
- Halliday, M.A.K., and Hasan, R. (1989). *Language, context, and text*. NY: OUP.
- Hasan, R. (1978). Text in the systemic-functional model. In W.U. Dressler (Ed.) *Current trends in textlinguistics. Research in text theory 2* (pp.228-246). Berlin and New York: Gruyter.
- Havelock, E.A. (1963). *Preface to Plato*. Cambridge, Mass: Harvard University Press.
- Havelock, E. A. (1976). *Origins of western literacy*. Toronto: The Ontario Institute for Studies in Education.
- Haviland, S. E., and Clark, H.H. (1974). What's new? Acquiring new information as a process of comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.
- Hayes, John R. (1996). A new framework for understanding cognition and affect in writing. In C.M. Levy and S.Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ.: Erlbaum.
- Heath, S. (1983). *Ways with words: Language, life and work in communities and classrooms*. London: Cambridge University Press.

- Hinds, J. (1983). Linguistics in written discourse in particular languages: Contrastive studies: English and Japanese. *Annual Review of Applied Linguistics*, 3, 78-84.
- Hinds, J. (1984). Retention of information using a Japanese style of presentation. *Studies in Linguistics* 8, 45-69.
- Hinds, J. (1987). Reader versus writer responsibility: A new typology. In Ulla Connor and R. B. Kaplan, *Writing across Languages: Analysis of L2 text* (pp. 141-152). Reading, MA: Addison-Wesley.
- Hoffman, Th. R. (1993). *Realms of meaning: An introduction to semantics*. New York: Longman House.
- Holmes, V.M. (1995). A crosslinguistic comparison of the production of utterances in discourse. *Cognition*, 54, 169-207.
- Hoover, L.M. (1992). Sentence processing strategies in Spanish and English. *Journal of Psycholinguistic Research*, 21 (4), 275-299.
- Hopper, P.J. and Thompson, S.A. (1980). Transitivity in grammar and discourse. *Language*, 56, 251-99.
- Hopper, P.J. and Thompson, S.A. (1984). The discourse basis for lexical categories in Universal Grammar. *Language*, 60 (4), 703-52.
- House, J.A. (1977). A Model for translation quality assessment. Tübingen, Germany: Gunter Narr Verlag.
- Hu, A., Brown, D., and Brown L. (1982). Some linguistic differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1, 39-40.

- Hunt, E., and Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, 98, 377-389.
- Hymes, D. (1972). Toward ethnographies of communication: The analysis of communicative events. In P. Giglioli(ED.), *Language and social context* (pp.21-43). Harmondsworth: Penguin. [excerpts from Hymes, D. (1966) Introduction: toward ethnographies of communication. *American Anthropologist*, 66 (6), 12-25].
- Jackendoff, R. (1994). *Language and human nature*. New York: Basic Books.
- Janda, Laura. (1996). Unpacking markedness. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.207-236). New York: Mouton de Gruyter.
- Johnson, Mark. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Johnstone, Barbara. (1987). Approaches to repetition. *Text*, 7 (3), 205-214.
- Jones, Larry B. (1983). *Pragmatic aspects of English text structure*. Arlington, TX: Summer Institute of Linguistics Publications in Linguistics, publication number 67.
- Just, M.A. and Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Just, M.A. and Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn & Bacon.
- Kail, M. (1989). Cue validity, cue cost, and processing types in sentence comprehension in French and Spanish. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic*

- study of sentence comprehension*, (pp. 77-117). Cambridge: Cambridge University Press.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16, 1-20.
- Kaplan, R. (1972). *The anatomy of rhetoric: Prolegomena to a functional theory of rhetoric*. Philadelphia: Center for Curriculum Development.
- Kaplan, R. (1988). Contrastive rhetoric and second language learning: Notes toward a theory of contrastive rhetoric. In A.C. Purves (Ed.), *Writing across languages and cultures. Issues in contrastive rhetoric* (pp.275-304). Newbury Park, CA: Sage.
- Kaplan, R. (1990). Concluding essay: On applied linguistics and discourse analysis. *Annual Review of Applied Linguistics*, 11, 199-204.
- Keenan, J.M., Baillet, S.D., and Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115-126.
- Kellogg, Ronald T. (1996). A model of working memory in writing. In C.M. Levy and S.Ransdell (Eds.) *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Mahwah, NJ.: Erlbaum.
- Kemppinen, A. (1988). Translating for popular literature with special reference to Harlequin Books and their Finnish translations. In S. Tirkkonen-Condit and S. Condit (Eds.), *Empirical studies in translation and linguistics* (pp. 113-14). Joensuu, Finland: University of Joensuu, Faculty of Arts.
- Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech*, 32 (1), 1-23.

- Kilborn. (1987). Sentence processing in a second language: seeking a performance definition of fluency. Unpublished doctoral thesis, University of California, San Diego.
- Kilborn, K., and Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence comprehension*, (pp. 257-291). Cambridge: Cambridge University Press.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W. (1979). On modeling comprehension. *Educational Psychologist*, 14, 3-14.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Kintsch, W. (1994). The psychology of discourse processing. In M.A.Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 139-160). New York: Academic Press.
- Kintsch, W. (1995). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In M.A.Gernsbacher and T.Givón (Eds.), *Coherence in spontaneous text* (pp.139-160). Philadelphia: J. Benjamins.
- Kintsch, W. and Greene E. (1978). The role of culture-specific schemata in the comprehension and recall of stories. *Discourse Processes*, 1, 1-13.
- Kintsch, W., and van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kintsch, W. and Vipond D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L.G. Nilsson (Ed.), *Perspectives of Memory Research* (pp. 325-66). Hillsdale, NJ: Erlbaum.

- Krulee, G.K., Fairweather P.G., and Berquist, S.R. (1979). Organizing factors in the comprehension and recall of connected discourse. *Journal of Psycholinguistic Research* 8, 141-163.
- Labov, William. (1973). The boundaries of words and their meanings. In Joshua Fishman (Ed.), *New ways of analyzing variation in English* (pp. 340-73). Washington, D.C.: Georgetown University Press.
- Labov, William. (1972). *Language in the inner city*. Philadelphia: University of Pennsylvania Press.
- Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*. Paperback edition. Chicago: University of Chicago Press.
- Lambrecht, Knud. (1987). On the status of SVO sentences in French discourse. In Russel Tomlin (Ed.), *Coherence and grounding in discourse* (pp.217-262). Amsterdam and Philadelphia: John Benjamins.
- Langacker, R.W. (1987). *Foundations of cognitive grammar* (Vol. 1, Theoretical prerequisites). Standord: Stanford University Press.
- Langacker, R. W. (1991). *Concept, image, and symbol: The cognitive basis of grammar*. New York: Mouton de Gruyter.
- Langacker, R. W. (1995). Possession and possessive constructions. In J.R. Taylor and R.Mac Laury (Eds.), *Language and the cognitive construal of the world* (pp.51-80). New York: Mouton de Gruyter.
- Langacker, R.W., and Casad, Eugene. (1991). Inside and outside Cora. In R.W. Langacker *Concept, image and symbol: The cognitive basis of grammar* (chapter II, pp. 33-55). New York: Mouton de Gruyter.

- Lautamatti, L. (1987). Observations in the development of the topic in simplified discourse. In U. Connor and R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 87-114). Reading, MA: Addison-Wesley.
- Leap. (1989). Written Ute English: Texture construction and point of view. *Journal of Navajo education* 2 (1), 3-9.
- Lee, Keedong. (1996). Getting at the meaning of make. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.389-423). New York: Mouton de Gruyter.
- Lindner, Susan. (1981). A Lexico-semantic analysis of verb-particle constructions with UP and OUT. Doctoral dissertation, University of California, San Diego. Available from the Indiana University Linguistics Club.
- Loftus, E. and Palmer, J.C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*. 13, 585-589.
- Lorch, R. (1995). Integration of topic information during reading. In R. Lorch and E. O'Brien (Eds.), *Sources of coherence in reading* (pp-279-294). Hillsdale, NJ: L. Erlaum.
- Lux, P. (1991). Discourse styles of Anglo and Latin American college student writers. Doctoral dissertation, Arizona State University.
- Lux, P., and Grabe, W. (1991). Multivariate approaches to contrastive rhetoric. *Lenguas Modernas* 18, 133-160.
- MacLaury, R. E. (1995). Vantage theory. In J.R. Taylor and R.Mac Laury (Eds.) *Language and the cognitive construal of the world* (pp. 231-276). New York: Mouton de Gruyter.

- MacWhinney, B. (1977). Starting points. *Language*, 53, 152-68.
- MacWhinney, B. (1987). *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127-150.
- MacWhinney, B., Pléh, Cs., and Bates, E. (1985a). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, 17, 178-209.
- Mandler, J.M., and Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 8, 11-151.
- McCawley, J.D. (1970). "Where do noun phrases come from?" In Roderick Jacobs and Peter Rosenbaum (Eds.), *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn.
- McClearly (1982). Transitivity and grounding in a Portuguese narrative text. Unpublished Master Thesis. University of California, Los Angeles.
- McCormack, P.D. (1977). Bilingual linguistic memory: The independence-interdependence issue revisited. In P.A. Hornby (Ed.), *Bilingualism: Psychological, social, and educational implications*, (pp.57-66). New York: Academic Press.
- McKay, S. (1989). Topic development and written discourse accent. In Donna Johnson and Duane Roen (Eds.), *Richness in writing: Empowering ESL students*. Longman: New York.
- McKoon, G., and Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior*, 20, 671-682.

- McKoon, G., and Ratcliff R. (1990). Textual inferences: Models and measures in D. A. Balota, G. B. Flores d'Arcais, and K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.
- McKoon, G., and Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466.
- McKoon, G., and Ratcliff, R. (1995). The minimalist Hypothesis: Directions for research. In C. Weaver, S. Mannes, and C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp.141-43). Hillsdale, NJ: L.Erlbaum.
- Menacere, M. (1992). Arabic discourse: Overcoming stylistic difficulties in translation. *Babel*, 38, 28-37.
- Meyer. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior* 7, 7-47.
- Meyer, D. E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Montano-Harmon, M. (1988). *Discourse features in the composition of Mexican, English as a second Language, Mexican-American/Chicano, and anglo high school students: Considerations for the formulation of educational policies*. Doctoral dissertation, University of Southern California, Los Angeles.
- Montano-Harmon, M. (1991). Discourse features of written Mexican Spanish: Current research in contrastive rhetoric and its implications. *Hispania*, 74, 417-425.

- Moragne e Silva, M. (1991). Cognitive, affective, social, and cultural aspects of composing in a first and second language: A case study of one adult writer. *Dissertation Abstracts International*, 52 (12), 4249^t (University Microfilms No.192-12, 595)
- Moxey, L.M., and Sanford, A.J. (1993). *Communicating quantities*. Hillsdale, NJ: Erlbaum.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: W.H.Freeman.
- Neubert, Albrecht. (1985). Text and translation. *Übersetzungswissenschaftliche Beiträge* 8, Leipzig: Enzyklopadie.
- Newmark, P. (1993). *Paragraphs on translation*. Clevedon, UK: Multilingual Matters Ltd.
- Nida. (1964). *Toward a science of translating*. Leiden, E. J. Brill:
- Noble, C.R. (1952). The role of stimulus meaning in serial verbal learning. *Journal of Experimental Psychology*, 43, 437-446.
- Nystrand, M. (1987). The role of context in written communication. In R.Horowitz and S.Jay Samuels (Eds.), *Comprehending oral and written language* (pp.197-212). New York: Academic Press.
- O'Grady W., Dobrovolsky M., and Aronoff M. (1993). *Contemporary linguistics: An introduction*. New York: St. Martin's Press.
- Oliveira de, Lúcia Pacheco. (1997b). Variação Intercultural na Escrita: Contrastes Multidimensionais em Inglês e Português [*Intercultural Variation in Writing*:

- Multidimensional Contrasts in English and Brazilian Portuguese*]. Unpublished doctoral dissertation, Pontificia Univerdade Católica de São Paulo, Brazil.
- Oliveira de, Lúcia Pacheco. (1997a). *Investigating discourse styles and finding cultural influences in writing*. Paper presented at the 31st. TESOL Convention, Orlando.
- Ostler, S. (1987). English in parallels: A comparison of English and Arabic prose. In U.Connor and R. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 169-185). Reading, MA: Addison-Wesley Publishing Company.
- Olson, D. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review* 47 (3), 257-281.
- Palinskar, A.S., and Brown, A.L. (1984). Reciprocal teaching of comprehension monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Perfetti, C. (1994). Psycholinguistics and reading ability. In M.A.Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp.849-894). New York: Academic Press.
- Pirolli, P.L., and Anderson, J.R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognitioin*, 11, 136-153.
- Posner, M.I., Snyder, C.R.R, and Davidson, B.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160-174.
- Potter, M. (1993). Very short-term conceptual memory. *Memory and Cognition*, 21 (2), 156-161.
- Potter, M.C., So, K.F., von Eckhardt, B., and Feldman, L.B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23-38.

- Prince, Ellen. (1981). Toward a taxonomy of given-new information. In P.Cole (Ed.), *Radical pragmatics*. New York: Academic Press, 223-56.
- Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W.C. Mann & S.A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text* (pp. 295-325). Amsterdam: John Benjamins.
- Radden, Günter. (1996). In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.423-458). New York: Mouton de Gruyter.
- Ratcliff, R.A., and McKoon, G. (1981). Does activation really spread? *Psychological Review*, 88, 454-462.
- Rayner, I, Carlson, M., and Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- Reed, Stephen. (1996). *Cognition*. Pacific Grove, CA: Brooks/Cole Publishing Company, (4th. edition).
- Reid, J. (1988). *Quantitative differences in English prose written by Arabic, Chinese, Spanish, and English Students*. Doctoral dissertation, Colorado State University.
- Reinhart. (1984). Principles of gestalt perception in the temporal organization of narrative texts. *Linguistics*, 22, 779-809.
- Reppen, R., and Grabe W. (1993). Spanish transfer effects in the English writing of elementary school children. *Lenguas Modernas*, 20, 113-128.
- Rosh, E. (1978). Principles of categorization. In E. Rosch and B.B. Lloyd (Eds.), *Cognition and categorization* (pp.27-48). Hillsdale: Lawrence Erlbaum.

- Rosh, E., and Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rubba, Johanna. (1996). The interaction of folk models and syntax: Case choice after prepositional verbs of cognition in German. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.837-856). New York: Mouton de Gruyter.
- Rumelhart, D.E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). Hillsdale, NJ: Erlbaum.
- Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- Rumelhart, D.E., and McClelland, J.L (Eds.). (1986). *Parallel distributed processing. Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Saliés, Tânia G. (1995a). *Recall Frequency: A pilot study*. Unpublished manuscript. Oklahoma State University, Stillwater.
- Saliés, Tânia G. (1995b). *ESL writing evaluation: The influence of discourse features*. Unpublished manuscript. Oklahoma State University, Stillwater.
- Saliés, Tânia G. (1995c). *Revisiting the role of L1 in L2 writing: Two case studies*. Paper presented at the 1995 OKTESOL Conference, Guthrie, OK.
- Saliés, Tânia G. (1995d). The role of L1 in L2: Perceptions of ESL students. Unpublished manuscript. Oklahoma State University, Stillwater.
- Samet, J. and Schank, R.C. (1984). Coherence and connectivity. *Linguistics and philosophy* (Vol VII, no.1, pp.57-82).

- Sanford, A.J., and Henderson, R.J. (1982). Anaphora resolution and episode signals. *Technical Report*, Department of Psychology, University of Glasgow.
- Scarcella, R. (1983). Discourse accent in second language performance. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 306-326). Rowley, MA: Newbury House.
- Schank, R.C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schreuder, R., and Flores d'Arcais, G.B. (1989). Psycholinguistic issues in the representation of meaning. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 409-436). Cambridge, MA: MIT Press.
- Schiffin, Deborah. (1994). *Approaches to Discourse*. Cambridge, MA: Blackwell.
- Scollon, R., and Scollon, S. (1990). Athabaskan English inter-ethnic communication. In D. Carbaugh (Ed.), *Cultural communication and intercultural contact*. Cambridge: Cambridge University Press.
- Scribner, S. and M. Cole . (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- Shen, Ya-Ming. (1996). The semantics of the Chinese verb "come." In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.507-540). New York: Mouton de Gruyter.
- Silva, T. (1992). L2 vs L2 writing: ESL graduate students' perceptions. *TESL Canada Journal* 10 (1), 27-48.
- Slobin, D., and Bever, T. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12, 229-265.

- Smith, E., and Swinney, D.. (1992). The role of schemas in reading text: A real-time examination. *Discourse Processes*, 15, 303-316.
- Snell-Hornby, M. (1988). *Translation Studies—An Integrated Approach*. Amsterdam: Benjamins.
- Sokolov, J.L. (1989). The development of role assignment in Hebrew. In B. MacWhinney, and E. Bates (Eds.), *The crosslinguistic study of sentence comprehension*, (pp. 118-157). Cambridge: Cambridge University Press.
- Sperber, D., and Wilson, D. (1986). *Relevance: communication and cognition*. Cambridge, MA: Harvard University Press; Oxford: Basil Blackwell.
- Sperber, D., and Wilson, D. (1995). *Relevance: communication and cognition*. Second Edition. Cambridge, MA: Harvard University Press; Oxford: Basil Blackwell.
- Sperber, D., and Wilson, D. (1991). Inference and Implicature. In Steven Davis (Ed.), *Pragmatics: A reader* (pp. 377-393). New York: Oxford University Press.
- Sridhar, S.N. (1989). Cognitive structures in language production: A crosslinguistic study. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp.209-224). Cambridge: Cambridge University Press
- Stein, N. L., and and Glen, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *Advances in discourse processes: Volume 2. New directions in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.
- Sumby, W.H. (1963). Word frequency and serial effects. *Journal of Verbal Learning and Verbal Behavior*, 1, 443-450.

- Swan, Michael, and Smith B (Eds.). (1987). *Learner English. A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.
- Sweetser, Eve. (1984). *Semantic structure and semantic change*. Ph.D. dissertation, University of California, Berkeley.
- Swinney, D. A. and Osterhout, L. (1990). Inference generation during auditory language comprehension. In A.C. Graesser and G.H. Bower (Eds.), *Inferences and text comprehension*. San Diego: Academic Press.
- Szwedek, A. (1984). Some problems of contrastive analysis and text linguistics. In J. Fisiak (Ed.), *Contrastive Linguistics Prospects and Problems*, (419-429). New York: Mouton Publishers.
- Tadros, A. (1994). Predictive categories in expository text. In M. Coulthard (Ed.), *Advances in written text analysis* (69-82). London: Routledge.
- Talmy, Leonard. (1978). Figure and ground in complex sentences. In J.H. Greenberg (Ed.), *Universals in human language*. Vol. IV: *Syntax* (pp. 625-649). Stanford, California: Stanford University Press.
- Tannen, Deborah. (1979). What's in frame surface evidence for underlying expectations. In R.O. Freedle (Ed.), *New Directions in Discourse Processing: Advances in Discourse Processes* Vol. 2 (pp. 137-181). Norwood, NJ: Ablex.
- Tannen, Deborah. (1982a). Oral and literate strategies in spoken and written narratives. *Language*, 58 (1), 1-21.
- Tannen, Deborah. (1982b). The oral/literate continuum in discourse. In D. Tannen (Ed.), *Spoken and written language: exploring orality and literacy* (pp.1-16). Norwood, N.J.: Ablex.

- Tannen, Deborah. (1989). *Talking voices: repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.
- Taraban, R., and Mc Clelland, J.L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectation. *Journal of Memory and Language* 27, 597-632
- Taylor, John R. (1987). Tense and metaphorisations of time in Zulu. In Lörcher, W. and Schulze, R. (Eds.), *Perspectives on language in performance. Studies in linguistics, literary criticism, and language teaching and learning*, 214-219. Tübingen: Narr.
- Taylor, John R. (1988). Contrasting prepositional categories: English and Italian. In Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 299-326). Amsterdam: John Benjamins.
- Taylor, John R. (1989). *Linguistic categorization: Prototypes in linguistic theory*. New York: Oxford University Press.
- Taylor, J. R. (1995). Introduction: On construing the world. In J.R. Taylor and R. MacLaury (Eds.), *Language and the cognitive construal of the world* (pp.1-22). New York: Mouton de Gruyter.
- Thompson, Sandra. (1987). "Subordination" and narrative event structure. In Russel Tomlin (Ed.), *Coherence and grounding in discourse* (pp.435-454). Amsterdam and Philadelphia: John Benjamins.
- Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9, 77-110.

- Thorndyke, P. W. (1979). Knowledge acquisition from newspaper stories. *Discourse Processes* 2, 95-112.
- Thorndyke, P. W. (1984) Applications of schema theory in cognitive research. In j. R. Anderson and S.M. Kosslyn (Eds.), *Tutorials in learning and memory*. San Francisco: W.H.Freeman.
- Tierney, R.J., and Mosenthal, J.H. (1982). The cohesion concept's relationship to the coherence of a text. *Technical Report no. 221*, Center for the Study of Reading. Champaign, IL: University of Illinois.
- Tirkkonen-Condit, S. (1989). Theory and methodology in translation research. In S. Tirkkonen-Condit and S. Condit (Eds.), *Empirical studies in translation and linguistics*, 17, pp. 3-18. Joensuu, Finland: University of Joensuu, Faculty of Arts.
- Tomlin, Russel. (1987). Linguistic reflections of cognitive events. In Russel Tomlin (Ed.), *Coherence and grounding in discourse* (pp.455-480). Amsterdam and Philadelphia: John Benjamins.
- Tomlin, Russel. (1995). Focal attention, voice, and word order. In P. Downing and M. Noonan (Eds.), *Word order in discourse*, (518-554). Philadelphia: John Benjamins.
- Toury, Gideon. (1991). What are descriptive studies into translation likely to yield apart from isolated descriptions? In K. Van Leuven-Zwart and T. Naaijken (Eds.), *Translation studies: The state of the art* (pp. 179-192). Amsterdam: Rodopi.
- Trabasso, T., and van den Broek (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24, 612-30.

- Trabasso, T., Secco, T., and van den Broek. (1984). Causal cohesion and story coherence. In H. Mandel, N.L. Stein, and T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83-111). Hillsdale, NJ: Erlbaum.
- Tyler, A., Jefferies, A., and Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, 7 (2), 101-110.
- Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly*, 26, 713-729.
- Van den Broeck, Raymond. (1978). The concept of equivalence in translation theory: Some critical reflections. In James Holmes, J. Lambert, and van den Broeck (Eds.), *Literature and translation*. Leuven, Belgium: Arco.
- Van Dijk, Teun. (1972). *Some Aspects of Text Grammars*. The Hague: Mouton.
- Van Dijk, Teun. (1988). *News as discourse*. Hillsdale, NJ: Erlbaum.
- Van Dijk, Teun, and Kintsch, Walter. (1978). Cognitive psychology and discourse: Retelling and summarizing stories. In W.U. Dressler (Ed.), *Current trends in text linguistics* (pp. 61-80). Berlin: Mouton de Gruyter.
- Van Dijk, Teun, and Kintsch, Walter. (1983). *Strategies in discourse comprehension*. New York: Academic.
- Verhagen, Ariel. (1996). Sequential conceptualization and linear order. In E.H. Casad (Ed.), *Cognitive linguistics in the redwoods* (pp. 793-820). New York: Mouton de Gruyter.
- Virtanen, Tuiya. (1990). On the definitions of text and discourse. *Folia Linguistica* 24, (no. 3-4), 447-455.

- Vonk, W., Hustinx, L., and Simons W.H. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes*, 7 (3-4), 301-33.
- Warren, R.M. (1970). Perceptual restorations of missing speech sounds. *Science* 167, 392-393.
- Watson, Rita, and Olson, David R. (1987). From meaning to definition: A literate bias on the structure of word meaning. In R. Horowitz and S. Jay Samuels, *Comprehending oral and written language* (pp.329-351). New York: Academic Press.
- Watters, K. J. (1996). Frames and semantics of applicative in Tepehua. In E.H.Casad (Ed.), *Cognitive linguistics in the redwoods* (pp.971-996). New York: Mouton de Gruyter.
- Weaver, C. and Kintsch, W. (1991). Expository text. In R. Barr, M. L. Kamil, P.B. Mosenthal, and P. D. Pearson (Eds.), *Handbook of reading research*, Vo. 2 (pp- 230-245). New York: Longman.
- Weise, Günter. (1988). Contrastive studies and the problem of equivalence in translation. In Fisiak, Jacek (Ed.) *Papers and studies in contrastive linguistics* (Vol 22, pp.188-195). The Polish-English Contrastive Project. (ERIC Document Reproduction Service No. ED 341 242).
- Whitney, P., Budd, D., Bramucci R., Crane, R. (1995). On babies, bath water, and schemata: a reconsideration of top-down processes in comprehension. *Discourse Processes*, 20, 135-166.
- Widdowson, H.G. (1979). *Explorations in applied linguistics*. London: Oxford University Press.

- Wierzbicka, A. (1989). Prototypes in semantics and pragmatics: Explicating attitudinal meanings in terms of prototypes. *Linguistics*, 27, 731-769.
- Wilson, Deirdre, and Sperber, Dan. (1991). Pragmatics and modularity. In Steven Davis (Ed.), *Pragmatics: A reader* (pp. 583-595). New York: Oxford University Press.
- Wilson, Deirdre, and Sperber, Dan. (1993). Linguistic form and relevance. *Lingua* 90, 1-25.
- Wills, Wolfram. (1982). *The Science of translation: Problems and methods*. Tübingen: Gunter Narr.
- Wills, Wolfram. (1988). *Kognition und Übersetzen: Zu Theorie und Praxis der menschlichen und der maschinellen Übersetzen* [Cognition and translation: Theory and practice of machine and human translation]. Tübingen: Max Niemeyer Verlag.
- Wills, Wolfram. (1989). Towards a Multi-facet Concept of Translation Behavior. *Target* 1 (2), 129-50.
- Winter, M.E. (1989). Diachronic prototype theory: on the evolution of the French subjunctive. *Linguistics*, 27, 703-730.
- Witte, S. and L. Faigley. (1981). Cohesion, coherence, and writing quality. *College Composition and Communication*, 32 (2), 189-204.
- Wixted, J.T., and Ebbesen, E.B. (1991). On the form of forgetting. *Psychological Science*, 2, 409-415.
- Woodley, Marie-Paule. (1987). *Non-nativeness in second language texts: The syntax factor*. Written language: British studies in applied linguistics 2. Papers from the Annual Meeting of the British Association for Applied Linguistics. (ERIC Document Reproduction services No. ED 346 745).

Yoshikawa, M. (1978). Some Japanese and American cultural characteristics. In M. H. Proesser (Ed.), *The cultural dialogue: An introduction to intercultural communication* (220-230). Boston: Houghton Mifflin.

APPENDIXES

APPENDIX A

GLOSSARY

1. **Base:** a specific cognitive domain or cognitive structures it presupposes (Langacker, 1991). For example, “food” functions as the base to define the category “rice” or “animal” functions as the base to define the category “tiger.”
2. **Category:** a class of related phenomena grouped together mentally on the basis of some kind of salient perceptual or functional similarity. It only exists within an individual’s cognitive structuring of his environment, and is either graded or not. For instance, the category *sem terra* (=landless) is the class of people who do not have a piece of land to harvest, raise cattle or live from in Brazil. That concept is well defined and may not make sense in countries like Switzerland, given that it is not part of the conceived reality. Other categories like “nouns” are graded with some members sharing most of the typical noun phrase attributes (e.g. “car”), the central members, and others displaying only a few of the attributes of the category (e.g. “there”), the more marginal members of the category.
3. **Category formation:** a functional cognitive process based on how the speech community adapts to the environment. The process includes the structuring of the conceived reality in familiar terms and the later integration with already existing ones to form a new structure, the category. It aims at providing maximum information with the least cognitive effort.
4. **Cognitive domains:** any sort of conceptualization that characterizes semantic structures as opposed to features or semantic markers. A cognitive domain may be a perceptual experience, a concept, a conceptual complex, an elaborate knowledge system.

5. **Concepts:** meaningful, motivated symbolic structures or mental representations of categories in the world. They function as recognition devices of instances of the category within an experientialist view of reason.
6. **Degree of prominence:** a general notion that may be realized syntactically, grammatically or prosodically and equated with salience in discourse. The easier it is to perceive a category the more prominent or salient it is. For example, words that show first in discourse/clause are more prominent than others. Elements that are explicitly mentioned in discourse through a full noun phrase are more prominent than those that need to be inferred. Entities that are stressed in spoken language are more prominent than those that have a weak accent. When a participant is the subject and the figure in the profiled relationship, it is the most prominent participant in the scene. Chafe (1995) shows that language gives more prominence to new ideas than to given ones, prominence being recognizable in terms of full nouns (more prominent) versus pronouns (less prominent) or ellipsis.
7. **e-site:** a schematic entity or sub-structure that figures saliently in the internal composition of a dependent structure and is put in correspondence with an autonomous structure. In addition, this sub-structure is elaborated by the autonomous structure and corresponds to it as a whole. For example, in the expression *large room*, [ROOM] is autonomous and elaborates the e-site of the dependent [LARGE] that profiles a relation between a bounded region and an abstract scale of comparison that determines size.
8. **Figure/ground:** Discourse is organized into figure/ground asymmetries. The participant that enjoys the status of figure, is the trajector, the initial or primary move

while the one that enjoys the status of ground is the landmark or the domain in which the relation trajector/landmark is staged.

9. **Folk models:** Conventional conception of certain concepts as opposed to the conception of experts. For example, geneticists and common people conception of “recessive genes.”
10. **Grounding:** the act of making information that is not accessible to the hearer accessible. When one grounds information one locates it with respect to the speech situation or makes it a dominion available to the addressees. That is what I am doing by writing a prologue for my study and defining terms my readers may not know (see Givón, 1990) for further details.
11. **ICM:** An Idealized Cognitive Model is a complex structure through which we organize our knowledge, generating category structures and prototype effects. It is not accessible to consciousness and is experientially-based. In one instance, Balinese and English people idealize their models of a “week” differently since the models do not exist objectively in nature but are created by people (see Lakoff, 1990). While the English week is a seven-day calendric cycle based on the movement of the sun, the Balenese is a complex combination of a lunar-solar calendar with ten different cycles.
12. **Image-schemas:** terminology used by Lakoff (1990) to refer to abstract conceptual structures that organize our experiences and that are configured as gestalts. They model perceptual invariants, having a rich internal logic and being highly predictive. The CONTAINER image-schema, for example, organizes a series of activities in terms of the structural elements “interior,” “boundary,” and “exterior” or simply in terms of in-out orientations. Constructions such as “pouring *out* coffee at breakfast,” “setting

out the dishes,” putting the toast *in* the toaster,” or spreading *out* the jam on the toast are construed based on the CONTAINER schema (examples from Johnson, 1987).

There is a container with an interior and the activities related to this container take *in* or *out* orientations. As such, image-schemas not only have structures of their own, but are used metaphorically to structure other concepts as in “that is out of question” or “he is in on the joke” (see Lakoff, 1990, pp.269-286 for a comprehensive discussion of image-schemas).

13. **Landmarks** (LM): entities that serve as reference point in a structure, like “car” in “the girl is behind the car.”
14. **Meaning**: conceptualization which is explicated in terms of cognitive processing. It includes experiences, feelings, abstract intellectual concepts and cultural and social awareness of the speech event.
15. **Metaphors**: descriptions of unfamiliar phenomena (target domain) in terms of already known phenomena (source domain) motivated by the structure of daily experience (see Lakoff, 1990 for details). For example, the Ph.D. is a JOURNEY metaphor correlates with the SOURCE-PATH-GOAL schema that is part of our everyday experience. The Ph.D. is the purpose or destination. To reach that destination, candidates need to take a series of actions that allow them to move from the starting point toward the degree. In other words, the metaphor experientially pairs the experience of a JOURNEY (known domain) with that of getting a Ph.D. (unknown domain). Within the cognitive approach, everything about language is a metaphor.

16. **Objectivity:** an aspect of perspective that constitutes the degree with which an entity is perceived in maximal contrast to the conceptualizer; the act of construing a given entity or situation as the object of conceptualization.
17. **Participant:** entities construed as subjects and objects in a clause; the actors on the stage.
18. **Profile:** an entity that can be defined with respect to a cognitive domain or base in a single two-dimensional representation. For example, in the phrase *past activities*, activities is the profile determinant or entity that determines the bounded region within the base that is under focus of assertion. In other words, the profile is a metaphor Langacker uses for the trajector or figure within a construction, the autonomous entity.
19. **Reification:** a process in which a state or event is converted into a referent. In Langacker's (1991) words, the incorporation of components of the verb schema [PROCESS] with components of the [NOUN] SCHEMA to form a complex high order constructional schema. That is the case of nominalizations in which a temporal process (= *to nationalize*) is converted into a referent (= *nationalization*).
20. **Setting:** a notion that depends on how a language user construes a situation. In general lines, it is the space in which participants interact and organize different scenes in discourse. Normally, it is expressed by an adverbial modifier such as geographical regions and extended time periods.
21. **Subjectivity:** a way of construing a conceived entity or situation (the object of conceptualization) as the perceiving individual himself (the conceptualizer). Such aspect of perspective is gradable. For example, "the sun" in "The sun sets down in the

West” is both the conceptualizer (in his role as agent) and the object of conceptualization.

22. *Trajectors* (TR): entities that are located in relation to a reference point, like “girl” in

The girl is behind the car.

TEXTS IN BRAZILIAN PORTUGUESE

Na Exploração, a Convivência Harmônica com a Natureza. (1994). *Meio Ambiente, Qualidade e Segurança*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

A exploração é o ponto de partida da indústria do petróleo. Sua missão é localizar as regiões mais favoráveis à ocorrência de petróleo no vasto território brasileiro, englobando a plataforma continental. Seu trabalho de campo -- representado basicamente por levantamentos sismográficos -- praticamente não interfere nas condições naturais, tanto em terra como no mar.

As equipes exploratórias passam vários dias em alto mar, nos áridos sertões ou mesmo nas densas florestas, expostas a ambientes hostis em regiões de difícil acesso, procurando fazer com que sua atividade nômade e efêmera seja imperceptível aos santuários de vida selvagem.

A presença da Petrobrás na prospecção de petróleo nas bacias sedimentares brasileiras é monitorada através de convênios com universidades e instituições científicas em vários estados, garantindo, desta forma, a obtenção de subsídios para que a prospecção de óleo e gás natural não interfira no ambiente natural.

Incentivo à Vida Natural. (1994). *Meio Ambiente, Qualidade e Segurança*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

Ao lado dos programas internos ligados à preservação ambiental, a Petrobrás vem desenvolvendo projetos de incentivo à vida natural, à preservação de espécies ameaçadas de extinção, ao esporte ao ar livre e à qualidade de bens e serviços.

Entre estas iniciativas, está o Projeto Tartaruga Marinha (Tamar). Coordenado pelo Instituto Brasileiro de Meio Ambiente e Recursos Naturais Renováveis (Ibama), este projeto -- apoiado pela Petrobrás -- luta contra a extinção de cinco espécies de tartarugas marinhas, entre elas a tartaruga-decouro, que pode atingir dois metros de comprimento e pesar 700 quilos. O Tamar protege a desova desses animais em quase todo o litoral brasileiro e já devolveu ao mar, em seus 11 anos de existência, mais de 1,5 milhão de filhotes de tartarugas.

Outra iniciativa nesta área é o programa "A Escola planta e colhe", responsável pela criação de cerca de 600 hortas escolares em vários pontos do País. Este projeto promove a integração com a comunidade, estimulando o amor pela natureza, a alimentação saudável e o trabalho como base do processo educativo.

A Petrobrás desenvolve, também, outros projetos que visam à melhoria da qualidade de vida e de produtos e serviços. Entre estas iniciativas estão a iniciação esportiva nas escolas e o Prêmio Petrobrás da Qualidade, conferido aos fornecedores que mais se destaquem na atividade de controle de qualidade.

Transportando com Segurança. (1994). *Meio Ambiente, Qualidade e Segurança*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

Para que o petróleo chegue às refinarias e os derivados ao consumidor, a Petrobrás opera extensa rede dutoviária e a maior frota de petroleiros do Hemisfério Sul. São utilizados navios petroleiros, vagões e

caminhões-tanque, uma extensa rede de dutos, que transportam os produtos até os pontos de armazenamento ou de abastecimento, e um conjunto de terminais marítimos.

A movimentação de combustíveis através destes meios exige rigorosas medidas de prevenção, que começam nos projetos dos novos empreendimentos e se estendem às operações.

Os terminais marítimos estão dotados de equipamentos especiais (barreiras de contenção, barcos recolhedores de óleo, coletores de óleo, etc.). Durante as operações de carregamento e descarregamento dos navios, realizadas por pessoal especializado e em constante treinamento, são adotadas severas medidas de prevenção e controle. Equipes de emergência estão sempre prontas para qualquer eventualidade.

Centros de combate à poluição por óleo foram instalados nos principais terminais marítimos, aumentando a segurança em caso de emergências. Em São Sebastião, litoral norte de São Paulo, foi construído um centro-modelo, que treina empregados da Petrobrás e de outras organizações para dar combate imediato e especializado a vazamentos acidentais de petróleo ou derivados.

Os dutos, além de passarem por inspeções frequentes, são dotados de dispositivos de segurança, como válvulas de bloqueio, que impedem a passagem de produtos em caso de anormalidade, resguardando as condições naturais das áreas marginais. Para maior garantia das comunidades, os dutos que ultrapassam áreas urbanas recebem, em seu projeto e construção, tratamento redobrado de prevenção, superando as próprias normas oficiais que regulam este tipo de empreendimento.

TEXTS IN ENGLISH TRANSLATION

Exploration in Tune with Nature. (1994). *Environment, Quality and Safety*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

Looking for petroleum is the starting point for the oil industry. One of Petrobrás' chief goals is to identify oil-prone regions within the vast boundaries of Brazil and in its territorial waters as well. Exploratory field work consists basically of seismographic surveys, which have practically no effect at all on natural conditions either on land or in the sea.

Petrobrás exploration crews spend long periods out in the ocean, in the Northeast's arid *sertão* regions, and even in Brazil's dense tropical forests. Exposed to inhospitable environments in remote places, these teams do their best to see that their nomadic, transient activity will leave no mark on any of wildlife's havens.

Exploratory work in Brazilian sedimentary basins is monitored under agreements with scientific institutions. This allows Petrobrás to carry out environmental projects and interact harmoniously with nature. In the Amazon, the Universidade Federal do Amazonas (UFA) and the Federal Environmental Protection Agency (INPA) are helping to research ways of preventing oil and natural-gas exploration from interfering with one of Brazil's most valuable natural resources -- the rain forest.

Transporting Safely. (1994). *Environment, Quality and Safety*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

Crude gets to the refineries and oil products to consumers via Petrobrás' extensive pipeline network and what is the largest tanker fleet in the Southern Hemisphere. Products travel from oil fields and sea

terminals to storage or supply posts via pipeline, ship, rail, and truck. Moving these fuels demands well-thought out safety measures and tough enforcement.

At Petrobrás' maritime terminals, specialized personnel with up-to-date training enforce strict compliance with prevention and control measures during onloading and offloading. Emergency crews are always on standby, ready to spring into action if needed.

Petrobrás has set up oil-spill control centers at its main sea terminals to better its emergency response capability. A model center has been constructed in São Sebastião, on the northern coast of São Paulo. There Petrobrás employees, together with professionals from other organizations, learn how to combat oil slicks rapidly and efficaciously.

Petrobrás routinely inspects its pipelines. These facilities are equipped with such safety devices as shut-off valves, which prevent inadvertent escape of products in case of abnormalities and thus protect the natural conditions of surrounding areas. Safety measures are doubled -- actually surpassing federal guidelines -- when pipelines are located near urban areas.

Lending Nature a Hand. *Environment, Quality and Safety*. Rio de Janeiro, RJ: Petróleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional

Alongside in-house programs in environmental preservation, Petrobrás sponsors external projects to protect endangered species and fosters respect for our natural habitat. Other projects encourage participation in open-air sports and commitment to quality goods and services.

An exciting example is the Tamar sea turtle project, coordinated by Brazil's Federal Environmental Protection Agency, Ibama, with Petrobras support. Tamar is fighting to save five species of sea turtles from extinction. One of these is the leatherback turtle, which can measure up to two meters and weigh as much as 700 kg (over 1500 lb.). The project protects the laying sites of these chelonians along 500 km (300 miles) of beaches in the state of Bahia. In Tamar's eleven years of existence, more than 400,000 baby turtles have been returned to the sea.

Under a program entitled *Planting and Harvesting at School*, some 400 gardens have been cultivated at public schools around the country. The project promotes community integration, instills a love of nature, and encourages healthy eating habits.

Together with the National Research and Technical Council (CNPq), Petrobrás sponsors the National Ecology Award, which honors new contributions in defense of the environment, in the form of essays, theses, and theories.

Other Petrobras-backed projects seek to improve Brazilians' quality of life as well as the quality of the goods and services they consume. These include a supervised beach-side gym program and sports activities at public schools. Additionally, suppliers with outstanding records in quality control are eligible to receive the Petrobras Quality Award.

TEXTS IN ENGLISH

Shell initiatives. (1993). *The Test of Tomorrow* (p.3). London, UK: Shell International Petroleum Company Ltd, Shell Centre.

The first written Shell policy statement on the environment was issued in 1969. Subsequent initiatives have reflected the need to improve environmental performance through proactive management and detailed programmes. The current policy calls for continuous improvement, with the ultimate aim—where possible—of eliminating emissions harmful to the environment.

Examples throughout exploration and production increasingly demonstrate the industry's ability and desire to change past practices, to implement new measures, and to protect and sustain the environment.

Shell E&P companies are already implementing a Safety Management System (SMS) as an extension of their existing Enhanced safety Management programme (described in the Shell Selected Paper, "How safe is safe enough?" by Richard Charlton). Operators and contractors should confirm that hazards have been systematically identified and assessed, and that there are arrangements to mitigate or control these hazards and deal with the consequences if all controls fail. The necessary information, training, auditing and improvement processes to achieve these objectives need to be in place.

Similar tools are available to tackle environmental issues, so it makes sense for environmental management to follow the SMS model and move to a formal Environmental Management System (EMS), and ultimately to a full Health, Safety and Environment (HSE) Management System. The way in which the environment should be integrated with economics in development decisions is a matter of judgment. However, the need to minimize the environmental impact of operations and to incorporate environmental values in decision-making is clear. Shell companies are committed to these goals. Some improvements can be, and are being, achieved quickly; others will take many years, major investment and new technologies.

The many good programmes already implemented in Shell operating companies to identify and manage wastes and potential impacts were formalized by the Environmental Management Guidelines, developed in 1986 and most recently updated in 1992. Their purpose is to provide direction on practical implementation of environmental management and conservation, with a focus on line-management involvement and responsibility. The objective is to maintain a healthy and sustainable environment.

From a starting point of full compliance with legislative requirements, Shell E&P companies pursue the environmental policy through company strategies and action. E&P personnel are involved in target-setting from the outset, and thus have full ownership of the activities and projects involved. Flexibility in the choice and application of technology to meet targets is a key factor.

Standards for industry are often established via legislation or regulation. While compliance with such standards is imperative, in some circumstances they could, or should, be exceeded—especially when a country does not have exacting environmental requirements, or when industry knowledge dictates a higher level of performance. Shell E&P companies recognize that working with others is also important. Organizations such as the E&P Forum, in consultation with the World Conservation Union (IUCN), are developing industry guidelines for operations in sensitive environments. Those for tropical forests were published in mid-1991, and those for mangrove swamps and arctic tundra were released in mid-1993. The International Chamber of Commerce, which has done much to promote environmental auditing and principles for sustainable development, is an inter-industry voice promoting environmental management.

Existing operations. (1993). *The Test of Tomorrow* (p.6). London, UK: Shell International Petroleum Company Ltd., Shell Centre.

Shell and E&P companies aim progressively to reduce emissions, effluents and discharges of waste materials that are known to have a negative effect on the environment, and where possible ultimately to eliminate them.

For an existing operation, the characteristics, level and impact of emissions should be known. An emissions inventory—updated annually—becomes a valuable database and is useful in identifying sources and prioritizing improvements. Experience has shown that the awareness created by the inventory triggers action to reduce waste, with as much as 50% coming from short-term measures such as leak prevention, segregation, reuse and better housekeeping. Determination of impact requires assessment of the fate of contaminants in the environment and understanding of their effects on the ecosystem. Combining emissions data with predicted or observed effect enables the operator to prioritize areas for emissions reduction. A reduction programme focuses on continually reducing the impact of harmful emissions through operational controls and monitoring, or design and implementation of improvements. The principal issues for the E&P industry are disposal of water produced with oil or gas; utilization of associated gas produced with oil; discharges of contaminated drilling waste; and management of other industrial waste, all of which must be done in compliance with all applicable company standards and regulatory requirements.

Texaco's 40th year with the Met. (1979). *Texaco 1979 Annual Report* (p. 29). White Plains, NY: Texaco Inc.

When the Metropolitan Opera in New York City opened its 1979-80 season in September, the occasion was carried nationwide on public television, made possible in part by a major grant from Texaco. The opera chosen for that first live color telecast of a Met Opening Night was Verdi's *Otello*, with tenor Placido Domingo, above, in the title role.

Three months later, on December 8, 1979, Texaco's sponsorship of the Saturday matinee broadcasts live from the Met began its 40th consecutive year. It represents the longest continuous sponsorship of a radio program by a single company in the history of broadcasting. Currently, these broadcasts are carried over the Texaco-Metropolitan Opera Radio Network, which includes over 300 stations in the United States and Canada.

Over the years, the Company's support of the Metropolitan Opera—particularly the Saturday matinee radio broadcasts—has remained a major corporate commitment. This commitment, in turn, has earned for Texaco praise from the press, as well as the loyalty of motorists and the gratitude of millions of opera fans.

Return of the Terns. (1975). *Exxon and the Environment* (p.106). Irving, Texas: Exxon Corporation. Shareholders Relations.

If you enjoyed seeing the swallows come back to Capistrano, you'll love watching the terns return to Exxon Company, USA's Grand Isle Gas Plant. Each April, right on schedule, thousands of the small, gull-like birds flock to this small island in the Gulf of Mexico. After flying 500 miles from the Yucatan Peninsula to Louisiana, the birds settle down within the gas plant confines to build nests and raise families.

Why Exxon's gas plant? Ornithologists say Exxon did the terns a good turn when the company brought in thousands of cubic yards of clam shells as a foundation for the maze of shiny towers which remove petroleum liquids from natural gas produced from nearby offshore platforms.

"The terns find an ideal nesting site among the clam shells," says Dr. George Lowery, Jr., professor of zoology at Louisiana State University and director of LSU's Museum of Natural Science. "They simply scratch out a small depression amid the shells, turn around a few times, and lay one to four speckled, cream-colored eggs. The birds and their eggs are perfectly camouflaged among the shells."

Gas plant employees welcome the migration each spring and do what they can to make the visitors feel at home. Burly oil workers spend hours locating the nests and marking them with stakes so as to warn away visitors. Vegetation goes uncut. Paths and walkways are rerouted to bypass concentrations of nests. Eggs laid in roadways are carefully moved to nests where foster mothers readily take them under their wing. Young chicks wandering onto roads are shooed back to the safety of the plant.

In September, as the days get shorter and the avian barometer indicates good flying weather, the terns prepare for their journey abroad. When the time is right, birds fill the fall sky as they head south to their wintering grounds across the Gulf.

Commenting on the phenomenon, a New Orleans outdoor writer says, "If there is any environmental conflict there, someone forgot to tell the terns about it, not to nest, that is, in the petroleum complex where they are bringing up their young by the thousands."

APPENDIX C

SAMPLE GLOSS OF A BRAZILIAN PORTUGUESE TEXT

Esforço para Preservar o Futuro
Effort (m) to safeguard the (m) future

Consciente de que meio ambiente, segurança industrial e qualidade
Certain (s.) that environment safety (f.) industrial and quality

são interdependentes, a Petrobras reuniu
are interdependent (pl) the (f.s.) Petrobras brought together (3rd.p.s; p.perf.)

a coordenação destas áreas num único
the coordination of these (f.pl.) areas (f.pl.) (under one) single (m.s.)

orgão, a Susema, ligado diretamente à
organization (m) the (f.s.) Susema, connected (m.s.) directly to the (s.f.)

Presidência, procurando integrá-las à
company chairman (s.f.) trying to integrate them (f.pl.) to the (f.s.)

missão principal da Companhia: produzir, comercializar
mission (f.s.) main of the (f.s.) company (f.s.): to produce, comercialize

e transportar petróleo e derivados para garantir
and distribute crude oil and oil products (m.pl.) to guarantee

o abastecimento do País.
the (m.s.) supplying (N; m.s.) of the (m.s.) country (m.s.)

Este esforço integrado, que pretende assegurar
This (m) effort (m.s.) integrated (m.s.) that intends to safeguard

a integridade do homem, compatibilizar a
the (f.s.) integrity (f.s.) of (m.s.) mankind (m.s.) reconcile the (f.s.)

produção... com o meio ambiente e
production (f.s.) of goods with the (m) environment (m) and

garantir a qualidade de produtos e serviços,
guarantee the (f.s.) quality (f.s.) of products and services

denomina-se Processo Petrobras de Meio Ambiente, Qualidade e Segurança.
calls itself Process Petrobras of Environment, Quality and Safety.

Petrobras reuniu a coordenação destas áreas...
Petrobras brought together the coordination of these areas

num único órgão... ligado diretamente à Presidência [da empresa]
under one single division answering directly to the chairman (of the company)

procurando [Petrobras] integrá-las à missão principal da Companhia...
trying [Petrobras] to integrate them to the mission main of the Company...

A Companhia busca se antecipar nas questões ambientais,
 The (f) Company endeavors itself to stay ahead in the (f.pl.) matters environmental

mantendo permanente diálogo com as comunidades afetadas
 keeping on going dialogue with the (f.pl.) communities (f.pl.) affected (f.pl.)

por seus empreendimentos.
 by its ventures (m.pl.)

Para melhor conduzi-los, a Petrobras realiza diagnósticos
 To better conduct them (m.pl.) the (f.s.) Petrobras undertakes diagnostics

e monitoramentos ambientais que,
 and monitoring (m.pl.) environmental (pl.) that

além de subsidiarem os Estudos de Impacto Ambiental (EIA),
 besides subsidize (3rd. p.pl.) the (m.pl.) Studies (m.pl.) of Impact Environmental (SIE)

resultam em benefícios para a própria sociedade,
 result (3rd.p.pl.) in benefits to the (f.s.) proper (f.s.) society (f.s.)

fornecendo novos dados sobre as condições ambientais
 providing new (m.pl.) data (m.pl.) about the (f.pl.) conditions environmental (f.pl.)

da região e dos ecossistemas envolvidos.
 of the (f.s.) region (f.s.) and of the (m.pl.) ecosystems (m.pl.) involved (m.pl.)

Esforço para Preservar o Futuro. (1994). *Meio Ambiente, Qualidade e Segurança*. Rio de Janeiro, RJ: Petroleo Brasileiro SA, Sercom - Serviço de Comunicação Institucional.

Consciente de que meio ambiente, segurança industrial e qualidade são interdependentes, a Petrobrás reuniu a coordenação destas áreas num único órgão, a Susema, ligado diretamente à Presidência, procurando integrá-las à missão principal da Companhia: produzir, comercializar e transportar petróleo e derivados para garantir o abastecimento do País.

Este esforço integrado, que pretende assegurar a integridade do homem, compatibilizar a produção com o meio ambiente e garantir a qualidade de produtos e serviços, denomina-se Processo Petrobrás de Meio Ambiente, *Qualidade e Segurança Industrial*. Um dos objetivos desse processo é atender às necessidades do presente sem comprometer o futuro.

A Companhia busca se antecipar nas questões ambientais, mantendo permanente diálogo com as comunidades afetadas por seus empreendimentos. Para melhor conduzi-los, a Petrobrás realiza diagnósticos e monitoramentos ambientais que, além de subsidiarem os Estudos de Impacto Ambiental (EIA), resultam em benefícios para a própria sociedade, fornecendo novos dados sobre as condições ambientais da região e dos ecossistemas envolvidos.

APPENDIX D

**SAMPLE OF BP TEXTS AND THEIR ENGLISH TRANSLATION
VARIANTS**

Esforço para Preservar o Futuro

Consciente de que meio ambiente, segurança industrial e qualidade são interdependentes, a Petrobrás reuniu a coordenação destas áreas num único órgão, a Susema, ligado diretamente à Presidência, procurando integrá-las à missão principal da Companhia: produzir, comercializar e transportar petróleo e derivados para garantir o abastecimento do País.

Este esforço integrado, que pretende assegurar a integridade do homem, compatibilizar a produção com o meio ambiente e garantir a qualidade de produtos e serviços, denomina-se Processo Petrobrás de Meio Ambiente, Qualidade e Segurança Industrial. Um dos objetivos desse processo é atender às necessidades do presente sem comprometer o futuro.

A Companhia busca se antecipar nas questões ambientais, mantendo permanente diálogo com as comunidades afetadas por seus empreendimentos. Para melhor conduzi-los, a Petrobrás realiza diagnósticos e monitoramentos ambientais que, além de subsidiarem os Estudos de Impacto Ambiental (EIA), resultam em benefícios para a própria sociedade, fornecendo novos dados sobre as condições ambientais da região e dos ecossistemas envolvidos.

Engenharia e Meio Ambiente

O Serviço de Engenharia da Petrobras incorpora em seus projetos, tanto na fase de estudos quanto na de execução, tecnologias ambientalmente viáveis, visando não apenas ao cumprimento da legislação em vigor, mas também a uma interação harmônica do homem com o ambiente.

A crescente integração da Petrobras com as universidades e instituições ambientais e a integração interdepartamental na Companhia visam esclarecer esses projetos às comunidades, desde a fase preliminar dos estudos até sua conclusão.

Striving to Safeguard our Future

Certain that environmental protection, industrial safety, and quality are interdependent, Petrobras has brought coordination of these areas together under one organizational roof within its corporate structure. This administrative unit, known as Susema, answers directly to the company chairman. Its main job is to ensure that this triad is an integral part of Petrobrás' efforts to guarantee domestic supply through the production, marketing, and transportation of crude and oil products.

Petrobras has drawn up a blueprint for safeguarding human lives, reconciling oil production with protection of the environment, and guaranteeing the quality of goods and services. A priority item on this Agenda for Environmental Protection, Quality, and Industrial Safety is meeting the needs of today's world without jeopardizing the world of tomorrow.

Petrobras endeavors to stay one step ahead in environmental matters by maintaining ongoing conversations with the communities affected by its ventures. In doing so, the company undertakes diagnostic surveys and monitors ecosystems. Findings not only subsidize environmental impact studies (EIS) but also benefit the society as a whole by furnishing new data on prevailing environmental conditions in these locations.

Engineering and the Environment

Petrobras' staff of engineers incorporates environmentally viable technology in its engineering design, from drawing board to project implementation. The goal is not only to comply with government regulations but also to promote nonaggressive interaction with the environment.

Through interdepartmental integration and growing cooperation with universities and environmental institutions, Petrobras clarifies its projects to the community during all stages of engineering.

Distribuição de Produtos com Qualidade e Segurança

A distribuição, realizada através de bases em vários pontos do País, depósitos em aeroportos, navios, caminhões e vagões-tanque e numerosa rede de postos revendedores, fecha o ciclo "do poço ao posto". } 1

A Petrobrás Distribuidora S.A. (BR), subsidiária que desempenha esta atividade, oferece ao mercado cerca de 350 itens, onde predominam combustíveis e lubrificantes que devem chegar a todas as regiões do Brasil. } 1

Com a responsabilidade de operar uma fábrica de lubrificantes e distribuir produtos para 7.000 postos e 8.000 clientes industriais, a BR emprega tecnologia avançada e rígidos critérios de qualidade, garantindo que o ciclo do petróleo se complete também com segurança e respeito ao meio ambiente.

Esta garantia é assegurada por produtos de qualidade e pela valorização profissional de seus empregados, que, em programas integrados de treinamento, recebem a formação necessária para tratar com a mesma ênfase a produção, a qualidade, a segurança e o meio ambiente. } 1

Secure Delivery of Top-rate Oil Products

The distribution process closes out the cycle from oil well to gas station. **The Petrobrás Group relies** on a network of supply bases located across the country, on airport storage facilities, on ship, highway, and rail transportation, as well as on a large number of retail outlets, to deliver its products to end-users. } 2

Petrobrás Distribuidora S.A., which goes by the trade name BR, is the group's subsidiary for oil-product distribution. **BR markets** some 350 items, mainly fuels and lubricants consumed in all corners of Brazil. } 2

In running its lubricant plant and distributing products to some 6,500 service stations and 7,000 industrial clients, BR employs advanced technology and rigid quality criteria, guaranteeing that the oil cycle is completed safely and at no undue risk to the environment.

This is possible thanks to top-grade products and a conscientious staff. **Through integrated training programs**, personnel learn to dedicate equal diligent to matters of production, quality, safety, and the environment. } 2

Environment, Quality and Safety. (1994). Rio de Janeiro, RJ: Petroleo Brasileiro AS, Sercom—Serviço de Comunicação Institucional.

21
VITA

Tânia Mara Gastão Saliés

Candidate for the Degree of Doctor of Philosophy

Thesis: TEXTS AS IMAGE-SCHEMAS: A CROSSLINGUISTIC ANALYSIS OF DISCOURSE IN THE LIGHT OF COGNITIVE LINGUISTICS

Major Field: English (Linguistics)

Biographical:

Personal Data: Born in Resende, RJ, Brazil, on April 11, 1959, the daughter of Waldyr and Lauda Gastão.

Education: Graduated from Colégio Estadual Marechal Souza Dantas High School, Resende, RJ, Brazil in December 1976; received certificate of proficiency in English language and literature from Centro de Cultura Anglo-Americana in December 1976; Bachelor of Arts degree in English with concentration in English-Portuguese translation and interpretation from Pontifícia Universidade Católica do Rio de Janeiro in December 1980; certificate of proficiency in German from the Goethe Institute in February 1981; certificate of proficiency in Spanish from Tulsa Junior College in October 1992; Master of Arts degree in English with concentration in TESL from Oklahoma State University in May 1994; completed the requirements for the Doctor of Philosophy degree with a major in Linguistics at Oklahoma State University in December 1997.

Experience: EFL teacher, Centro de Cultura Anglo-Americana, Resende, Brazil, March 1976, to March 1977; literary and technical translator, Record Serviços de Imprensa SA, Rio de Janeiro, May 1980 to May 1981; simultaneous translator; bilingual executive secretary, Xerox do Brasil SA, June 1981 to February 1990; Thomas De La Rue SA, February 1990 to September 1991; ESL volunteer teacher, Robert Grove Elementary School, YWCA, Tulsa, Oklahoma, October 1991 to January 1994; Research Assistant for the Mid-level Writing Assessment Project, Oklahoma State University, Stillwater; Faculty Member, English Department, Pontifícia Universidade Católica do Rio de Janeiro; Faculdade da Cidade, March 1996 to present.

Professional Memberships: TESOL, OKTESOL, BRAZTESOL

Honors: Tulsa Junior College Presidents' Honor Hall
University Center at Tulsa's Honor Hall, Phi-Kappa Phi

Awards: TESOL Prentice Hall Regents Larry Anger Fellowship for Graduate Study
Oklahoma State University Humanities Graduate Research Excellence