Syntax in the ST Register: Effect on Writers' Choices and Readers' Comprehension

Judith D. Banks Strother
Syntax in the ST Register:

Effect on Writers' Choices and Readers' Comprehension

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, Prof. ir. M. Tels, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op vrijdag 9 februari 1990 te 16.00 uur

door

JUDITH DIANNE BANKS STROTHER

geboren te Corpus Christi, Texas, U.S.A.
Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. D. Wim Vaags
en
Prof. dr. Thomas N. Huckin

Copromotor:

Dr. Jan M. Ulijn
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>x</td>
</tr>
<tr>
<td>I. READING AND WRITING IN THE SCIENCE AND TECHNOLOGY REGISTER</td>
<td></td>
</tr>
<tr>
<td>1.1 The Reading-Writing Connection</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The ST Register</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Definition of the ST Register</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Development of the ST Register</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3 Rhetorical Functions Within the ST Register</td>
<td>5</td>
</tr>
<tr>
<td>1.2.4 The Syntax of the ST Register</td>
<td>6</td>
</tr>
<tr>
<td>1.3 The Effect of ST Syntax on Reading and Writing</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Readability Formulas</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2 ST Syntactic Elements Thought to Cause Difficulty</td>
<td>11</td>
</tr>
<tr>
<td>1.3.3 The Call for Clear Writing to Improve Readability</td>
<td>14</td>
</tr>
<tr>
<td>1.4 The Role of Background Knowledge on Reading and Writing</td>
<td>15</td>
</tr>
<tr>
<td>1.5 The Purpose of the Research in Examining the Reading-Writing Connection</td>
<td>17</td>
</tr>
<tr>
<td>II. THE EFFECT OF A WRITER'S BACKGROUND KNOWLEDGE AND LANGUAGE KNOWLEDGE ON CHOICE OF SYNTACTIC STRUCTURES</td>
<td>21</td>
</tr>
<tr>
<td>2.1 Psycholinguistics and Writing</td>
<td>21</td>
</tr>
</tbody>
</table>
4.3 The Effect of Time in the Reading and Writing Tests  
4.4 General Conclusions  
4.5 The Interactive Reader-Writer Readability Model  
4.6 Practical Implications and Suggestions for Further Research  
4.7 Teaching Implications  

Appendix A  
Appendix B  
Appendix C  
Appendix D  
REFERENCES  
SAMENVATTING
LIST OF TABLES

CHAPTER II

2.1 Subjects of FIT/TUE Writing/Reading Studies ........................................... 28
2.2 Syntax Use in Computer Science Context for FIT/TUE 1 .............................. 33
2.3 Results for Syntax Use by American (L1) Subjects for FIT/TUE 3 ............... 40
2.4 Results for Syntax Use by Dutch (L2) Subjects for FIT/TUE 3 ................. 41
2.5 Subjects Writing Scientific and Nonscientific Paragraphs .......................... 46
2.6 Mean Number of Sentences Containing ST Syntactic Structures within Scientific and Nonscientific Paragraphs .................................................. 49

CHAPTER III

3.1 Subjects of FIT/TUE Writing/Reading Studies ........................................... 60
3.2 Mean Reading Comprehension Scores of a Computer Science Text for FIT/TUE 1 ...... 61
3.3 Mean Times (in minutes) to Read the Text and Complete the Test for FIT/TUE 1 ... 61
3.4 Mean Reading Comprehension Scores of Computer Science and Fiction Texts for FIT/TUE 2 .......................................................... 64
3.5 Mean Times (in minutes) to Read the Text and Complete the Test for FIT/TUE 2 ... 64
3.6 Results of Reading Comprehension Scores for FIT/TUE 2 ............................ 66
3.7 Results for Reading Comprehension Scores by Dutch (L2) Subjects for FIT/TUE 3 ... 69
3.8 Results for Reading Comprehension Scores by American (L1) Subjects for FIT/TUE 3 ............................................................... 70
3.9 ANOVA Results for Subjects Reading a Computer Science Text .................. 75
3.10 Mean Number of Correct Answers for L1 and L2 Subjects Reading a Computer Science Text 77

3.11 Mean Times (in minutes) for L1 and L2 Subjects Reading a Computer Science Text 78

3.12 Number and Distribution of Subjects Reading a Fiction Text 82

3.13 Multiple Regression Results for Experts (Scientists) vs. Nonexperts (Nonscientists) 83

3.14 Mean Reading Comprehension Scores and Reading Times for the Fiction Text 85

CHAPTER IV

4.1 Summary of Results of the Effect of Specific Variables on a Writer's Choice of Syntax 88

4.2 Summary of Results of the Effect of Syntax on Reading Comprehension Scores 93
# LIST OF FIGURES

## CHAPTER I

1.1 Elements of Readability 16
1.2 An Interactive Reader-Writer Readability Model 19

## CHAPTER II

2.1 A Cognitive Model of the Writer 23
2.2 Cognitive Processing Models for Reading and Writing 24
2.3 Syntax of Texts Used in FIT/TUE Experiments 27
2.4 Syntax Use in Computer Science and Fiction Texts for FIT/TUE 2 36
2.5 Syntax Use in CS vs. Fiction Texts by L1 and L2 Subjects for FIT/TUE 3 42
2.6 Effect of Language Background, Linguistic Context, and Background Knowledge on Syntax Use for FIT/TUE 3 44
2.7 Interaction of Language Background and Background Knowledge on Syntax Use in Scientific and Non-Scientific Paragraphs 50

## CHAPTER III

3.1 Cognitive Processing Models for Reading and Writing 55
3.2 A Cognitive Model of the Reader 56
3.3 Effect of Linguistic Context on Reading Comprehension Scores for Fiction Texts for FIT/TUE 3 68
3.4 Effect of Linguistic Context on Mean Times to Complete the Test for FIT/TUE 3 71

## CHAPTER IV

4.1 An Interactive Reader-Writer Readability Model 101
ABSTRACT

Writing and reading are considered as two indispensable parts of one written communication chain. Writers must write text which is to be comprehended by specific readers, and conversely, readers must be able to comprehend what writers produce. Syntax is certainly a factor in the way texts are written and the way and the ease with which they are read. This book examines some questions that relate to the effect of syntax on both reading and writing specifically within the ST (science and technology) register.

Is there a difference in the way scientific and technical texts are written as compared with texts on a nonscientific subject? If a difference exists, is it a function of the writer--e.g., cognitive system, background knowledge of the subject being written about (expert vs. nonexpert in the text topic), or language knowledge (native or first language, L1, vs. nonnative or second language, L2)? Is it a function of the text itself--the subject matter being written about? Is it a function of the reader or the audience for the text--specialists in the topic field vs. lay readers? Or, is the difference the result of an interaction between the writer's capabilities and the features of the text to be produced?

Is there a difference in the way ST texts are read compared with CL (common language) texts? If a difference exists, is it a function of the reader--e.g., cognitive system, background knowledge of the subject being read, language knowledge--or is it a function of the text--e.g., syntax, lexis, discourse structure? Or, is the difference the result of an interaction between the reader's capabilities and the features of the text to be read?

This book provides at least limited answers to the above questions, as well as examines the interrelationship between the reading and writing processes as they relate to ST materials.

For the writing studies, the primary questions dealt with what causes a certain writer to choose particular syntactic structures within a
certain topic or semantic context. The overall conclusion to be reached is that all subjects--L1 and L2, expert and nonexpert--prefer to use ST syntax in scientific texts and CL syntax in nonscientific texts. The issue of the influence of any surrounding cotext (the sets of all sentences surrounding a given sentence in a text) on syntactic choices was eliminated for most subjects. These studies give strong evidence that writers have an equal option of choosing either the ST or the CL register in syntactic structuring, yet the topic or semantic context itself is the primary factor in determining a writer's choice of syntax, regardless of that writer's language background or expertise in the subject matter being written. Thus, at least for the syntactic element of the ST register, that register is a psycholinguistic reality. For the ST register, the topic of the text has a stronger impact on ST syntax use than background knowledge or language knowledge. The text's subject automatically suggests the syntax to be used regardless of the writer's background knowledge and language knowledge.

Since writers tend to tune their linguistic register to the topic of the text, independent of background knowledge (expert vs. nonexpert) and language knowledge (L1 vs. L2), technical writing is, for both natives and nonnatives, not only a matter of language and specialist knowledge training, but also of gaining insight into the psycholinguistic aspects of writing.

For reading, the primary question was whether the ST register, with its supposedly more difficult structures, such as passives, nominalizations, participles, and infinitives, is more difficult to comprehend than the CL register. These reading studies strongly concluded that the syntax of the ST register did not cause more comprehension difficulties than that of the CL register for any of the subject groups. In addition, both L1 and L2 readers do not seem to have their comprehension greatly hampered by violated syntax (sentence elements presented in alphabetical rather than natural order). The great extent to which sentences with a violated syntactic structure can be understood by both
native and nonnative readers gives additional evidence for a conceptually guided and partially parallel strategy for processing texts.

The practical implications for technical readers and writers and for teachers of technical reading and writing for both native and nonnative learners are presented.
CHAPTER I
READING AND WRITING IN THE
SCIENCE AND TECHNOLOGY REGISTER

1.1 The Reading-Writing Connection

It is time researchers in the discipline called English bridge the gulf between the reading researcher and the writing researcher. There are now many trained writing researchers who can collaborate with the trained researchers in reading, for the act of writing is inseparable from the act of reading. You can read without writing, but you can't write without reading (Murray, 1982, p. 141).

Tierney and Pearson (1983) have concisely summarized the fact that reading and writing are inextricably intertwined by saying that if one is to truly understand the reading-writing connection, one must begin to view reading and writing as essentially similar processes of meaning construction.

Both are acts of composing. From a reader's perspective, meaning is created as a reader uses his background of experience together with the author's cues to come to grips both with what the writer is getting him to do or think and what the reader decides and creates for himself. As a writer writes, she uses her own background of experience to generate ideas and, in order to produce a text which is considerate to her idealized reader, filters these drafts through her judgments about what her reader's background of experience will be, what she wants to say, and what she wants to get the reader to think or do. In a sense, both reader and writer must adapt to their perceptions about their partner in negotiating what a text means.

Rankin (1988, p. 120) added that "if writing can be considered an intertwining set of decisions made with regard to the choice and arrangement of ideas and words in a text, then reading is the other end of the
exchange: a set of decisions leading to an interpretation of the meaning expressed by the writer, based on textual cues."

This book looks at the reading-writing connection within a rather narrow scope, that of the ST (Science and Technology) register. The research is further narrowed to examine the effect of the ST register on writers' syntactic choices and on readers' comprehension of text and the interrelationship of the reader, the writer, and the text on the issue of readability.

1.2 The ST Register

1.2.1 Definition of the ST Register

There is an increasing recognition of the general field of Languages for Specific Purposes (LSP) and with that there is the increasing need to clearly define those areas which exist as subsets of LSP. In this book, the language for specific purposes is English (ESP), and using Mackay's (1975, 1978) rationale, ESP generally refers to the learning and/or teaching of a particular language for a clearly utilitarian purpose, with the teaching of English not an end in itself, but an essential means to a goal which is clearly identified. Thus, the purpose for which the learner is studying is specific, not the language itself.

English for Science and Technology (EST), which is the main focus of this book, is one of the major subdivisions of ESP. Trimble (1985, p. 6) defined EST by saying that it "covers that area of written English that extends from the 'peer' writing of scientists and technically oriented professionals to the writing aimed at skilled technicians." In between are several types. Peer writing is exemplified by books and articles written by experts in one field for other experts in the same field or for experts in a related field. Technicians are those who differ from experts only in that they sometimes lack equivalent training in theory. "In sum, EST covers the areas of English written for academic and professional purposes and
of English written for occupations (and vocational purposes), including the often informally written discourse found in trade journals and in scientific and technical materials written for the layman" (Trimble, 1985, p. 6). Thus, the idea of audience is a crucial one for the writer to consider when producing materials within the ST register.

EST may thus have an occupational focus, included in English for Occupational Purposes (EOP) for those who are already working in fields such as engineering, physics, or computer science, or it may have an academic focus, English for Academic Purposes (EAP) for those students at universities or colleges studying scientific and/or technical subjects. This book deals primarily with EAP since all subjects of the reported research studies were university students.

The terms ESP and EST imply a special language to some extent; however, while EST has specific lexical and syntactic characteristics, it is certainly not a separate language. Register analysis has been used by researchers to attempt to demarcate the constraints of EST, but those boundaries are far from clear.

According to Halliday and Hasan (1989, p. 12), a register is a semantic concept which can be defined as "a configuration of meanings that are typically associated with a particular situational configuration [social context] of field [nature of social action], mode [symbolic organization, function of text, rhetorical category], and tenor [nature and interrelationships of participants]. But since it is a configuration of meanings, a register must also, of course, include the expressions, the lexico-grammatical and phonological features that typically accompany or realize these meanings." Although register is considered by some to be a semantic concept, its linguistic features must also be examined to provide a complete picture.
1.2.2 Development of the ST Register

When analyzing the development of the concept of an ST register, one could start with the work of Peter Strevens (Halliday, McIntosh, & Strevens, 1964; Halliday, McIntosh & Strevens, 1973), Jack Ewer (Ewer & Latorre, 1969), and John Swales (1971). Their goal was to identify the grammatical and lexical features of a specific register. They found little distinctive in the sentence grammar of EST beyond a tendency to favor particular forms: present tense, passive voice, and nominal compounds. Others, such as Godman and Payne (1981, p. 23) concluded that the "grammar in scientific statements does not appear to differ from that in the general language, although the interpretation of some grammatical structures is different from, and in some cases more precise than, the interpretation considered normal in the general language." All of this, of course, reinforces EST as a subset of general language.

These researchers, and others who emphasized register analysis, had a pedagogic aim of making ESP courses more relevant to students' needs and of producing a syllabus which gave high priority to language forms which students would meet in their scientific studies.

The second phase of development of register analysis focused on the field of discourse or rhetorical analysis. This emphasized the identification of organization patterns in text and specified the linguistic means by which these patterns are signalled (Hutchinson & Waters, 1987, p. 11; see also Bley-Vroman, 1978). This also called attention to descriptions of expository versus narrative text for scientific and technical purposes.

With the understanding of the above difficulties, it is still valuable to limit a study within a register, following Swales' (1978) suggestion that specific areas of study (specifically statistical surveys of term types, etc.) within a register help at least provide a framework for pedagogical design.
1.2.3 Rhetorical Functions Within the ST Register

As Widdowson (1981) points out, the identification of a register involves making statements about its formal properties as a type of English text. The other way, to describe it as rhetoric, involves making statements about the English textualization of a type of discourse—the way a particular language realizes the concepts and functions of a particular type of discourse. Textualization is concerned with the relationship between linguistic form and rhetorical function in written language. "In a more restricted sense it can be defined as the process of describing how grammatical structures match up with meaning, or in Widdowson's words, 'how they express elements of discourse'" (Weissberg & Buker, 1978, p. 322).

Rhetoric, which is one part of discourse, is therefore the process a writer uses to produce a certain piece of text. "This process is basically one of choosing and organizing information for a specific set of purposes and a specific set of readers. An EST text is concerned only with the presentation of facts, hypotheses, and similar types of information. It is not concerned with the forms of written English that editorialize, express emotions or emotionally based argument or are fictional or poetic in nature" (Trimble, 1985, p. 10). Thus, exposition, expressing rhetorical functions such as reporting, explaining, and evaluating, is one of the characteristics of the ST register. In expository text, there are several ways to express these rhetorical functions common to the ST register, e.g., prepositions, nouns, subordinate clauses introduced by conjunctions or participles, and infinitive constructions with or without conjunctions (Ulijn, 1984, p. 72). A technical writer must therefore consciously choose certain syntactic structures for the specific rhetorical functions of a scientific or technical text within a particular context.

The notion of register proposes a very intimate relationship of text
to context: "indeed, so intimate is that relationship, it is asserted, that the one can only be interpreted by reference to the other. Meaning is realized in language (in the form of text) which is thus shaped or patterned in response to the context of the situation in which it is used" (Halliday & Hasan, 1989, p. vii). Text, therefore, is defined as language that is functional, language that is doing a particular job in some context. Although text is written in the form of words and structures, it really is in and of itself a semantic unit. Thus, text is both product and process. It is product in the sense it is output, having certain constructions which can be represented in systematic terms. It is process in the sense of the fact that it represents a continuous process of semantic choice. Halliday and Hasan (1989) go on to assert that function is a fundamental property of text, that it is basic to the evaluation of the semantic system. "This amounts to saying that the organization of every natural language is to be explained in terms of a functional theory" (p. 17).

1.2.4 The Syntax of the ST Register

According to the above discussion, a register incorporates a subset of the total language, especially in grammatical and lexical selections, which varies according to use, depending on the activity or context in which the language is functioning. To thoroughly analyze the ST register, one would have to look at many parameters, including morphology, syntax, and lexis. However, for the purposes of this book, the analysis is first narrowed to the syntax of the ST register and then is further narrowed to four syntactic structures that are the most common in the ST register--nominalizations, passives, participles, and infinitives.

There have been an increasing number of studies on the characteristics of the EST register (see, e.g., Swales, 1971; Todd Trimble et al., 1978; Tarone et al., 1981; Bazerman, 1984; Ulijn, 1979, 1984; Tessman, 1985; Trimble, 1985). When comparing the results of these
research studies, one can find conclusions that are at times conflicting. Most studies (e.g., Todd-Trimble, et al., 1978; Ulijn, 1979, 1984; Trimble, 1985) show that some syntactic structures, specifically, nominalizations, passives, participles, and infinitives, appear with greater frequency in scientific and technical (ST) text than in text which is written in common language (CL), providing the rationale for the selection of these syntactic structures for use in the studies in this book. For example, Swales (1971, p. 39) points out that "it is probable that in any physics, chemistry or engineering textbook, at least one-third of all the finite verbs will be in the passive. . . . Therefore, it is clear that scientists and engineers use the passive much more frequently than most other kinds of writers."

Some authors (e.g., Wingard, 1981; Tarone, et al., 1981) caution that, while it is safe to generalize that in most EST texts the passive, infinitive, participle, and nominalization are used with greater frequency than in common language, care must be taken not to overgeneralize. There exists the possibility that grammatical features may occur with varying frequency from text to text, even within a single field of study. It has been shown, for example, by Tessman (1985) that scientific writing is not stable in its style requirements. Questionnaires and interviews among researchers of eight national research laboratories in the US show evidence that not all science writers avoid first person or insist upon using the passive voice. On the other hand, Bazerman (1984), in his analysis of the style used throughout the history of the Physical Review, discovered that while syntactic complexity and the use of ST structures have remained constant, other factors are variable. For example, there has been a significant increase in the use of nominalizations, abstraction of sentence subjects, and complexification of the multiword noun phrase.

There seem to be some solid reasons for choosing certain syntactic structures within the expository text of the ST register. Making
the actor secondary to the action is essential to the ST register. Nominalizations, participles, and passives are simply stylistic variations that do so. In some cases, their use is the result of a writer's attempt to remain impersonal. On the other hand, the nature of the subject matter itself seems to dictate the use of the passive. Stress on the action, not on the person doing the acting, is inherent in science and technology, and passives are thus used to stress the action.

Tannebaum and Williams (1968) suggested that there seems to be a psycholinguistic distinction between the active and passive deriving specifically from the characteristics of the particular encoding situation rather than from the nature of the language code itself. They also concluded that the usage of the two forms differs, where either form may have been appropriate, "particularly in terms of the apparent conceptual foci of the situations. Under such an interpretation, it is assumed that the passive form, for example, was more apt to be encountered where the attention was directed on the "natural" object of the situation. Such a consistent contextual feature may then become a distinctive cue associated with the use of the passive voice. When the demand characteristics of a given encoding situation are similar in this respect—e.g., when attention is directed at the acted-upon object. . . this cue may be activated, and the passive-voice form rises in the hierarchy of possible response forms" (Tannenbaum & Williams, 1968, p. 250). This would support the rationale of passives being used to stress the action rather than the actor in scientific text. At the same time, it introduces a research question which is raised in the current studies: Is the ST register a psycholinguistic reality within the writer/reader?

1.3 The Effect of ST Syntax on Reading and Writing

The concept of readability incorporates many aspects of text, e.g., lexis, text linguistics, cohesion, paragraph length and organization,
concreteness or imageability, referents, and reader motivation. However, the current research is limited to the syntactic aspects of readability. One important question of this research is whether the structures which occur with greater frequency in most ST texts--passives, nominalizations, infinitives, and participles--cause the reader or writer more difficulty than do other structures. The answer to that question should, at least for these particular syntactic features, help in determining what makes a text more readable or more comprehensible to a reader, or conversely, what text elements reduce readability or make a text less comprehensible. While other researchers have attempted simply to describe the syntactic elements of the ST register, the current study seeks to contribute to the understanding of the effect of that register's syntax on the reader-writer-text interaction.

1.3.1 Readability Formulas

Historically, researchers have addressed the concept of readability and have made numerous and varied attempts to measure it; yet, the results have been quite mixed. For the past 30 or 40 years, much attention has been paid to formulas that claim to measure readability. Most of these formulas (e.g., Flesch, 1948, 1949; Dale & Chall, 1948; Fry, 1958) only take into account surface elements in a text such as sentence length (a syntactic feature) and word length. In addition to lack of analysis of the effect of specific syntactic structures, these readability formulas have a number of other shortcomings, especially when used within the ST register.

Since readability formulas are designed to analyze prose text, problems arise when these formulas are applied to scientific, technical, and mathematical materials which contain a great deal of numerical and symbolic language and specialist vocabulary in addition to prose (Hartly, 1985).
There now seems to be consistent research evidence against using readability formulas as guidelines for technical writers to write and edit technical and legal documents for adult readers (Duffy, 1985; Felkner, Redish, and Peterson, 1985, p. 56). Selzer (1983, p. 75) says, while it is doubtful that readability formulas can predict the readability of technical texts after they are written, it is important to note that these formulas "cannot be used to guide the production of readable technical documents. Because readability formulas are often based on sentence length and word length, because sentence length and word length may correlate with readability, some people have mistakenly assumed that longer sentences and words cause reading difficulty. Flesch himself was one who confused correlation and causation." Psychologist David Pearson (1974, p. 160) warns about the "common error in interpreting correlational data by assuming that correlation means causality. The fact that sentence length, sentence complexity, or any other factor correlates with . . . difficulty. . . does not imply that altering these correlates will reduce difficulty."

When Duffy and Kabance (1982) tested the effects of rewriting prose passages to readability guidelines, they concluded that rewriting to improve the readability score of a passage will not necessarily produce more comprehensible text. Unlike the lexical domain, where major features of a given register (terminology or "jargon") tend to be more obviously difficult for nonspecialists, not all syntactic features typical of that register are automatically difficult.

To be effective, readability measures should include careful manipulation of cohesive structures of a text, as suggested by Olsen and Johnson (1990). Coherence based textual revision may increase the difficulty of passages as indexed by traditional readability formulas, but can enhance comprehension, as has been evidenced by Bech, et al. (1984) for second grade material.
Since readability formulas do not adequately describe features of a text that influence comprehension, new approaches to textlinguistics, or the structure of text beyond the sentence, have been developed (Olsen & Johnson, 1990; Binkley, 1988). Selzer (1983) stresses the need to analyze factors such as the effect of topic sentences, given vs. new information, and propositions (analogous to the "kernel sentences" of transformational grammar. While these systems have contributed to a general understanding of the reading process, they do not specifically address the contribution of individual syntactic features to the readability of a particular text.

1.3.2 ST Syntactic Elements Thought to Cause Difficulty

According to some research, there are specific syntactic structures known to cause difficulty in comprehension. Coleman (1962) examined three syntactic simplification strategies. On the one hand, breaking a compound sentence joined by and into two sentences had no effect. Raising clause fragments (e.g., participle, gerund, and infinitive phrases) in a complex sentence to the status of a full sentence resulted in only marginal improvement in comprehension. On the other hand, breaking sentences joined by coordinate conjunctions other than and caused reliably better comprehension. Bhatia (1984) ascertained some syntactic problems in British legislative writing. While the English judicial register might have a different syntactic impact on reading than the EST register would, aside from specialized terminology, many of the features are the same. Law Professor Robert Benson (1985) humorously summarized the difficulties with the judicial register, only two of which are addressed by readability formulas:

There exist scores of empirical studies showing that most of the linguistic features found in legalese cause comprehension difficulties. Legalese is characterized by
passive verbs, impersonability, nominalizations, long sentences [which imply also participle and infinitive constructions], idea stuffed sentences, difficult words, double negatives, illogical order, poor headings, and poor typeface and graphic layout. Each of these features alone is known to work against clear understanding.

Some syntactic variables are referred to in many guidelines which are designed to help writers produce readable material. Technical writing textbook authors such as Rew (1989) ask editors to be sure most sentences are in the active voice and to justify carefully any passives used. While Rew stresses that the use of the passive has some advantages, e.g., emphasizing the receiver of the action, making a statement impersonal, or softening responsibility, she concludes her advice to technical writers by stating that "when you want to be crisp and clear, you will use the active voice most of the time--for its shorter and more forceful sentences. Only in special circumstances will you use passive constructions. Wrongly used, passive constructions can confuse the readers, and overused, they can numb the brain" (Rew, p. 190).

There is the suggestion that the use of nominalizations (converting verbs into nouns) makes sentences more complex because they are indirect. (Dawkins, 1975; Harrison, 1980, p. 23; Klare, 1980, 1985; Price, 1984, p. 143, American Psychological Association, 1983). While most of the writers' handbooks are aimed at L1 technical writers, Huckin and Olsen (1983) address these same complexities for nonnative technical writers.

Thus, authors of textbooks and handbooks such as those mentioned above tell writers to increase readability by keeping sentences active, thereby avoiding passives and nominalizations, and by making separate sentences of dependent clauses, thereby avoiding participle and infinitive constructions, but do these directives truly increase readability?
Original texts are often thought to be more difficult to comprehend than simplified ones. One reason for this is the complexity of the syntax in original texts. Certainly, a knowledge of syntax is a requirement for second language reading comprehension, especially at less advanced levels. Nilagupta (1977), Alderson and Alvarez (1978), and Berman (1984) present some evidence on that score. There is, of course, a syntactic threshold which must be overcome when one starts learning to read in a foreign language as Alderson (1984) suggests and Stone’s (1985) experimental results confirm. Cooper’s (1984) results suggest that first year university students in Malaysia were insecure in their understanding of meaning carried by tense, aspect, modality, nonfinite participial clauses and sometimes hypothetical conditions and passivization in English text.

At advanced levels, however, according to most current experimental literature, syntax plays only a minor role in both native and second language reading comprehension. Garrod (1984), for example, reports that the lowest factor contributing to variation in reading time in English as a native language (E1) was syntactic complexity. In reading EST materials in English (E2), Venezuelan science students experience fewer syntactic than lexical difficulties (Akirov & Salager, 1985). This is in line with the results of Israeli science and technology students whose syntactic knowledge seems to have contributed the least to their English (as L2) reading comprehension (Weiss, 1985). There is a higher correlation between reading achievement and the recognition of individual lexical items than between reading achievement and knowledge of grammar, speaking fluency, or any other linguistic skill (Saville-Troike, 1979). The limited importance of syntax for advanced readers is strongly evidenced by Ulijn (1981), but those experiments dealt with French instructions written in a common rather than a scientific register.
1.3.3 The Call for Clear Writing to Improve Readability

There is a definite need to analyze the features of text that enhance readability and then to incorporate those features. One of the many reasons for such research has been to answer the call for simple clear language in both the US and Britain. For the purposes of this book, the resulting question is: what is the effect of syntax on the definition of simple clear language?

In 1978, President Carter sent Executive Order 12044 to all government departments mandating the use of plain English in government documents. Many states have followed this trend, and while Plain Language Laws differ from state to state, their basic goal is to simplify all consumer documents and contracts. In addition, some court cases have used readability measurements as pivotal instruments in cases where it was determined that the language of the document in question (e.g., directions on a package or Medicare instructions) was too difficult for the defendant to comprehend (Battison, 1981; Fry, 1988, pp. 78, 88-89). In Britain, there is also a Plain English Campaign, with its headquarters in Stockport. That organization publishes newssheets entitled *Plain English* and runs training courses, advises on and conducts research in all aspects of plain English.

Because of this emphasis on plain English in the US and Britain, the answer to the above question for both L1 and L2 readers might be of particular importance to technical writers who are to design highly communicative specialist and lay documentation, for example, in computer science and information technology. The answer is also relevant to EST teachers who wish to adapt original specialist and lay materials for second language learners.

There has been a call for experimental research, for example by Davis and Kantor (1982), to determine if syntax really affects readability
as the proponents of readability formulas claim. (See the one by Flesch, 1949, and others referred to by Harrison, 1980). Therefore, one of the purposes of this book is to examine the syntactic features of the ST register, the effects of the register on the grammatical choices of the writer as he or she realizes meanings, and the effect these choices have on the reader's comprehension of what is written.

1.4 The Role of Background Knowledge on Reading and Writing

A less frequently studied element of readability is the reader's background knowledge of the subject being written about related to schemata (see the extensive research by Bransford, et al. on this, e.g., Bransford, Stein, & Shelton, 1984; Britton & Tesser, 1982; and Carrell, 1983). Readers usually recall more of what they read when the information presented is tied to their prior knowledge. In one study of domain-related text (text content of the subject matter domain in which knowledge appears), it was found that subjects with background knowledge of the text being read processed domain-related text more efficiently than did subjects without such background (Fincher-Kiefer, Post, Greene, & Voss, 1988). The difficulty of the lexis is automatically an element of this factor.

There is considerable variation in the way readers use their background knowledge, depending on the purpose they have in mind. Physicists reading physics literature, for instance, predominantly present a dynamic interplay between the purpose they have in mind in reading a text and the schema of the background knowledge they use to comprehend that text (Bazerman, 1985).

Kieras (1980), using L1 subjects, found that readers can form macro-representations of the text, even when they do not have background knowledge of the central concepts of the text and even when
surface cues mislead them. They can make very general decisions, for example, about what technical terms mean and can then obtain comprehension based on those generalizations. Schmalhofer (1982) examined the abilities of expert and nonexpert (in computer science) readers to comprehend a manual about a new computer programming language, LISP. He found that none of the subject groups showed any memory for the surface characteristics of sentences, which indicates that the subjects did not process individual sentences by surface features, but rather used general comprehension strategies for encoding the text. Background knowledge had a significant effect on what they learned from the text. "In summary, non-schema (low knowledge) subjects learned the text, whereas schema (high knowledge) subjects, in addition, learned the technical skill of LISP programming from the text" (p. iv).

The predominant role of background knowledge in foreign language reading has been ascertained in a number of experimental reading studies, mostly in EST contexts, such as Ulijn (1975, which dealt with technical French); Mohammed and Swales (1984); Osman (1984); Alderson and Urquhart (1985); Moy Yin (1985); Levine and Haus (1985, which was on Spanish); and Peretz and Shoham (1990).

Comprehension is influenced by the reader’s decoding ability and text topic familiarity. Even a skilled reader will have difficulty comprehending a relatively simple text when the topic is totally unfamiliar. In addition, research in the cognitive approach to human information processing has shown that whenever the reading level of the material changes, the nature of the cognitive processing changes too (Kintsch & Van Dijk, 1978; Kintsch & Vipond, 1979; Kintsch & Miller, 1984; Zakaluk & Samuels, 1988).

Present cognitive psychology and psycholinguistics emphasize schema-based, conceptually guided, interactive, and partially parallel strategies to process texts. (A good description of the current schema
theory is found in Rumelhart, 1981, and Carrell, 1987). This implies that higher order elements, such as text structure and content words, allow a reader to build up a conceptual representation of a text and that surface elements such as syntactic structures which are less sensitive to that representation are simply overlooked.

Since readers often use different schemas or routines than writers intend, this has to be considered in writing readable texts. Huckin (1983) points out that, using protocol analysis, cognitive scientists have discovered that experts process information differently than novices. If writing for experts, cues should be included that trigger, in long-term memory, the experience and knowledge needed to understand the text. If writing for nonexperts, analogies from common knowledge can be used to help the reader acquire new knowledge which eventually becomes part of his or her own expert systems as new knowledge accumulates and is called forth in a variety of applications. The basic message of cognitive research is that if the thought-processing patterns to be used by the reader are identified, the writer can design the text to facilitate that kind of processing (Samuels, 1988).

1.5 The Purpose of the Research in Examining the Reading-Writing Interaction

The need to further investigate the effects of language knowledge (L1 vs. L2), background knowledge (expert vs. nonexpert), and text topic (scientific vs. nonscientific) on syntactic choices in LSP writing and to examine the effect of these elements and their interaction with syntactic complexity of a text on reading comprehension has been established.

In this book, only the syntactic element of readability is focused on. An attempt is made to describe the interrelationship of the reader and the writer with the readability of a particular text. In many of the studies reported earlier, a readability index was the result of a linguistic
description of the text and not a cognitive analysis of how the reader processes text. The current research seeks a psycholinguistic description of the role of syntax in readability.

Basically, then, readability provides the key link between the reader and the writer. Figure 1.1 presents a very simplified model of the elements to be considered in readability.

![Diagram of Readability Elements]

Fig. 1.1 Elements of Readability

The readability model presented in Figure 1.2 is an expansion of Figure 1.1 and is used to describe how the reader, writer and text interact and what the influencing factors on each are. In general, reading does not occur without a written text, and text does not exist without the writer's production of that text. This research includes the writer's psycholinguistic interaction with the text topic (a nonlinguistic element) and thus moves the concept of readability toward the writer instead of focusing it exclusively.
Figure 1.2 An Interactive Reader-Writer Readability Model
on the reader and the text itself. The influencing factors of background knowledge (expert vs. nonexpert in a given field) and language knowledge (L1 vs. L2) are discussed for both the writer (Chapter II) and the reader (Chapter III). A detailed description of the model is presented in Chapter IV after research results are given.

Thus, this book and the experiments reported herein may make contributions in several important areas:

1. It should make a contribution to psycholinguistic theory as it moves toward further understanding of the writing process. It examines the effect of a writer's background knowledge and language knowledge on his or her syntactic choices and also examines the effect of the text topic and the immediate context in which the writer is working on those syntactic choices.

2. It should make a contribution to psycholinguistic theory as it moves toward further understanding of the reading process. It examines the effect of these same variables on a reader's comprehension of text and the role of syntax in that comprehension process.

3. Since writers must produce text which is to be read and readers must attempt to comprehend what writers have produced, it is vital to examine both activities and the interrelationship between them. It is critical to determine what elements do or do not give a reader difficulty so technical writers can more appropriately structure their text to enhance readers' comprehension of that text. These studies, therefore, attempt to examine the contribution of a writer's use of syntax to readability.

4. In LSP pedagogical situations, the latest empirical theories and results should be known and applied so that teachers in the field can help future technical writers and technical readers develop their skills in the most efficient way.

5. In industrial situations, these empirical results should help LSP professionals write for maximum readability.
CHAPTER II
THE EFFECT OF A WRITER'S BACKGROUND KNOWLEDGE AND LANGUAGE KNOWLEDGE ON CHOICE OF SYNTACTIC STRUCTURES

2.1 Psycholinguistics and Writing

This book takes a psycholinguistic approach to the syntactic element of writing in an attempt to examine how a writer determines which form of syntax to choose within a particular text topic or context and what other factors, such as language background and knowledge of the subject to be written are influencing factors. The text topics or domains chosen for this book are in the field of science and technology and they are contrasted with nonscientific topics to analyze whether text topic is also an influencing factor on a writer's choice of syntax.

There are various models which have been proposed for demonstrating the cognitive processes used by a writer. One of the newest attempts at modeling has been made by connectionism which stresses the interactions among the various parts of the brain (see, e.g., Schneider, 1987). Rather than a controlling mechanism in the brain such as a monitor or overlord, in connectionist models, the brain is seen as a self-organized system with a central controlling or processing system being unnecessary because the representations (activation patterns in the neural networks) compete with each other. The most adequate activity is selected and controls the system. Thus, there is a great deal of parallel activation and interactivity (Vaags, 1989).

However, there is still a great deal of reservation about such models because (among other reasons) of the limited amount of empirical evidence dealing with human cognitive systems (Schneider, 1987). For the purposes of this book, therefore, a more conservative psycholinguistic model of writing is used. This model allows for three cognitive strategies:
using serial, strictly parallel, or partially parallel processes for conceptualizing (thinking) and formulating (transforming thoughts into written text) (Ulijn & Gobits, 1986; Ulijn, 1987; Ulijn, Goetschalck, & Schouwstra, 1989). First, it is important to take a brief look at the model for the writer's method of converting information from concept to written text. As can be seen in Figure 2.1, Ulijn's (1987) model of the writer's internal process of producing sentence level text is used. This parallels the work discussed in this book since this research is limited to sentence level syntactic analysis and does not delve into such elements as lexis and rhetorical strategies.

In Ulijn's model of the writer, the overlord, in a constant interaction among all systems and with feedback and anticipation circuits, supervises all mechanisms. For the writer to produce a sentence, the text and sentence generator, working within a specific theme, retrieves certain conceptual fragments from the conceptual system, attempts to formulate them into sentences by retrieving the appropriate words from the lexicon, and then provides them with suitable syntactic structures. As Ulijn (1987, p. 10) points out, "the syntax and lexis of the language to be written in determine which semantic dimensions are going to make up the conceptual structure of a sentence." The syntactic structuring, the conceptualizing and the lexicalization are operating in an interactive partially parallel way. Then, the script producer chooses the appropriate script (e.g., ideogram or alphabet), splitting the output up into letters, words, and blank spaces to produce the sentence.

In both reading and writing, there is constant interaction between form and meaning. Since reading in its most primitive stage relates form to meaning and writing relates meaning to form, one has to deal with both and with the interrelationship between the two. Figure 2.2 shows a comparison among serial, parallel, and partial parallel processing models for both reading and writing. Indirectly, some syntactic aspects of the
Figure 2.1: A Cognitive Model of the Writer

The Operation of Text & Sentence Parser in Reading

<table>
<thead>
<tr>
<th>The Operation of Text &amp; Sentence Generator in Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Staging</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

1. Serial

Text → SA → LA → CA

or

CA → LA → SA

Bottom Up

or

S → L → C → Text

Top Down

2. Parallel

CA

LA

SA

Text

Interactive

3. Partial Parallel

Text → SA → LA → SA

Superficial

CA

Thorough

Revision

Interactive

Figure 2.2 Cognitive Processing Models for Reading and Writing

Legend:

C = conceptual retrieval
CA = conceptual analysis
L = lexicalization
LA = lexical analysis
S = syntactic structuring
SA = syntactic analysis

Source: Ulijn 1987

24
partial parallel processing model are tested in the current research, so it has been adopted here for illustrative purposes.

As is discussed in Chapter III for reading, a thorough syntactic analysis is not necessary; however, for writing, it cannot be avoided since the writer must choose particular syntactic structures to express meaning. Therefore, it is the purpose of this chapter to try to determine what causes a writer to make syntactic choices within a given context.

There are four writing studies reported in this chapter. The first three were cooperative efforts between Florida Institute of Technology (FIT) in the United States and Eindhoven University of Technology (TUE) in The Netherlands (Ulijn & Strother, 1987; Strother & Ulijn, 1989a; Strother & Ulijn, 1989b). The fourth was the first of two independent studies completed for the book. (The second is a reading study and is discussed in Chapter III.)

The writing studies reported herein all attempt to determine why a certain writer chooses a particular syntactic form within a particular text topic or context. While these studies examine the same basic issue, they are progressively more refined in eliminating influences from the surrounding linguistic cotext on a writer's syntactic choices. The definition and the rationale for the use of the term linguistic cotext (rather than linguistic context) is given by Van Dijk (1972):

... the term cotext is used to denote the sets of all sentences following and preceding a given sentence in a text, instead of the traditional but ambiguous term '(verbal) context,' which is often used for cotext as defined above, but also for texts (or rather discourses) preceding or following (or overlapping) a given discourse. We will reserve the term 'context' for the set of all relevant factors of the communication process (or speech act) which are not properties of the text/discourse itself.

Thus, when the term cotext is used herein, it refers to the formal syntactic
structure of the sentences surrounding a particular piece of text, not to the semantic context related to the topic of the text.

The two specifically delineated kinds of cotext used in the studies in this work have been defined in terms of registers; namely, ST (Science/Technology) linguistic cotext and CL (Common Language) linguistic cotext. As discussed in Chapter I, the ST register contains a larger number of syntactic structures such as nominalizations, passives, participles, and infinitives than does the CL register, which is characterized by active constructions replacing these structures. That provides the rationale for selecting those four structures to be the focus of all of the reading/writing studies discussed in this book.

2.2 General Methodology for the FIT/TUE Studies

To make the results of the studies comparable, the same two texts were used in the first three writing studies (FIT/TUE 1, FIT/TUE 2, and FIT/TUE 3) and the same criteria for selecting L1 (English as a first or native language) and L2 (English as a second language) expert and nonexpert subjects were used. Each test was pretested before being administered to subject groups. The two text topics provided a strong contrast in context. The ST selection was from Miller’s (1982), "Mass storage systems and evolution of data center architectures," from the journal Computer. The CL selection was from Jack London’s (1964) fictional work, "To Build a Fire." The syntactic structuring of each of the texts for each of the studies is summarized in Figure 2.3 and is explained in detail in each of the study descriptions below.

The criteria for selecting subjects included a selection for language background (L1 vs L2) and for background subject knowledge of the text, in this case, computer science. In general, the L1 subjects were American university students and the L2 subjects were mixed ESL students at American universities and/or Dutch students at Dutch
universities. The mixed ESL subjects were students at Florida Institute of Technology and had an advanced knowledge of English with TOEFL (Test of English as a Foreign Language) scores in the 500-575 range. The Dutch subjects have English as a second or foreign language with about six years of English study at the secondary school level and considerable exposure to English textbooks and review articles. While we do not have TOEFL scores for this particular group of Dutch subjects, Educational Testing Service (1987) data gives an average TOEFL score of 603 for Dutch students for the period 1978-1986. It can be assumed that this score is representative of the Dutch subjects used in these studies since their
educational level is comparable to the EST test group.

To screen for background knowledge of computer science, subjects were considered expert if they were at least in their third or fourth year of study in a computer science major. Nonexperts had no formal training in computer science and were usually humanities majors. A summary of the subjects for the three studies is given in Table 2.1 and the subject selection is explained in detail for each study.

Table 2.1

Subjects of FIT/TUE Writing/Reading Studies

<table>
<thead>
<tr>
<th>Background Knowledge</th>
<th>FIT/TUE1</th>
<th>FIT/TUE 2</th>
<th>FIT/TUE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Language Knowledge</td>
<td>Language Knowledge</td>
<td>Language Knowledge</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>L2</td>
<td>L1</td>
</tr>
<tr>
<td>Experts</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Nonexperts</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>n = 96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For each of the studies, the ST text or both the ST and CL texts were used. Within each text, certain sentences were selected for the subjects to complete. Subjects were given the words, in alphabetical order, needed to complete each sentence (violated syntax). In each of these sentences, it was possible to make either CL or ST syntactic choices. The instructions told them they could: change the order of the groups of words; change the forms of words (for example, adjective into adverb or verb into noun); conjugate the verb; and add some words if necessary to complete the sentence. Examples were given, and, while care was taken to be sure that subjects understood the task, no specific feedback was given. The following are examples from each of the texts:

Example 1 comes from the computer science text:

(at higher system level) (current devices with optical data disks) (replace) (...)(...). The programmer must develop considerably more software to take advantage of optical data disks.

Possible solutions are:

(a) Replacement of current devices with optical data disks will probably occur at a higher system level. The programmer must develop considerably more software to take advantage of optical data disks.

(b) If, at a higher system level, current devices are replaced with optical data disks, the programmer must develop considerably more software to take advantage of optical data disks.

(c) To replace current devices with optical data disks at higher system levels, the programmer must develop considerably more software to take advantage of optical data disks.

(d) We can replace current devices with optical data disks at a higher system level. The programmer must develop considerably more software to take advantage of optical data disks.
(e) At a higher system level, current devices with optical data disks often replace many tapes. The programmer must develop considerably more software to take advantage of optical data disks.

(f) The programmer must develop considerably more software to take advantage of optical data disks when he replaces current devices with optical data disks at high system levels.

Example 2 is from the fiction text:

He was a warm-whiskered man, but (the hair on his face) (his high cheekbones and nose) (not) (protect).

Possible solutions are:

(a) He was a warm-whiskered man, but there was no protection of his high cheekbones and nose by the hair on his face.

(b) He was a warm-whiskered man, but the protecting of his high cheekbones and nose was not done by the hair on his face.

(c) He was a warm-whiskered man, but his high cheekbones and nose were not protected by the hair on his face.

(d) He was a warm-whiskered man, but the hair on his face was not able to protect his high cheekbones and nose.

(e) He was a warm-whiskered man, but the hair on his face could not protect his high cheekbones and nose.

(e) He was a warm-whiskered man, but the hair on his face did not protect his high cheekbones and nose.

In all of the writing studies, the criteria for operationalizing the distinction between ST and CL syntactic structuring were determined in the following way. Two independent raters were asked to evaluate the student solutions by analyzing the syntactic structure of the words between parentheses and then to mark the solutions according to these directions:
1. If a sentence contains a nominalization, a participle, an infinitive, or a passive, mark it $S$ ( = ST syntax).
2. If not, mark it $C$ ( = CL syntax).
3. If the sentence cannot be analyzed, i.e., has no answer, or has an answer that does not include the elements asked for or does not make sense, mark it $O$.

There were few differences in opinion between the two raters, but those were discussed and led to a definitive mark; therefore, interrater reliability was very high. Subject answers were analyzed to decide logic problems. A logic problem occurred when the subject wrote a perfectly grammatical sentence which was classified as being in the ST or CL register, but which changed the meaning from that originally given in the article. For example, in one sentence, which should have read, "... the problem is separating and clustering sets of files and providing a mechanism. ...," a few subjects wrote "... the problem is separating clustered sets of files. ..."--still in the ST register, but different in meaning from the original. This kind of response was tabulated as a logic problem.

In Example 1 from the ST text given above, the underlined fragments (a), (b), and (c) are examples of $S$ and (d), (e), and (f) are examples of $C$. In this particular sentence, the interpretation of the verb "replace" is crucial. If the subject tries to make a nominalization, a passive, or a participle of "replace," the score is $S$. If the subject tries to conjugate "replace" to make an active construction, the score is $C$. In the CL text, the same is true for the interpretation of the verb "protect." Thus, in Example 2 from the CL text given above, the underlined fragments (a), (b), (c), and (d) are examples of $S$ and (e) and (f) are examples of $C$.

Kies (1985) correctly states that stylistic variants between CL and ST registers are often due to functional differences, which explains, for instance, the use of active or passive. In the studies reported in this book, however, the two solutions $S$ or $C$ are equally possible in English with
hardly any difference in function.

As was mentioned in Chapter I, there is a strong connection between writing and reading; therefore, in each of the writing studies (FIT/TUE 1, 2, and 3), there was also a reading comprehension test on the text(s) being used. Times to complete the task were also recorded. These results are presented in full in Chapter III.

The question, therefore, is what causes a writer to use a particular syntactic structure within a particular text? Will someone who writes about scientific or technical (ST) subjects choose the syntactic structures which characterize scientific text? Do second language students prefer the supposedly less difficult common language syntax? Is scientific background knowledge a significant factor in making syntactic choices?

2.3 FIT/TUE 1--Writing Within an ST Text (ST Cotext)

The purpose of the first study (FIT/TUE 1) was to examine the commonalities and differences between L1 and L2 syntactic structuring in LSP writing for those writers who have a strong background in the subject being written about and those who do not. (See Strother & Ulijn, 1989a, for a full description of this study.)

The subjects of this study were 48 American students at Florida Institute of Technology and Valencia Community College in Florida and 48 Dutch students at Eindhoven University of Technology and Eindhoven College of Business Administration in The Netherlands.

The text used for this study was the authentic version of the computer science journal article which had a predominance of ST syntactic structures. Within the text, ten sentences were selected for the subjects to complete. A 2 x 2 ANOVA (two way, with repeated measures) was used with the following independent variables and conditions:
1. Background Knowledge:  
(1) Computer Science (experts)  
(2) Humanities (nonexperts)

2. Language Knowledge  
(1) Native English (American)  
(2) ESL (Dutch)

The hypothesis that writers with an expert knowledge in a scientific field choose significantly more ST syntactic structures than nonexpert writers in that field and the hypothesis that background knowledge contributes significantly more to syntactic structuring in writing than language knowledge were supported. A third hypothesis, that L1 writers choose significantly more ST syntactic structures than L2 writers was not supported. Table 2.2 summarizes the use of syntactic structures within the ST text for both L1 and L2 subjects.

Table 2.2
Syntax Use in Computer Science Context for FIT/TUE 1

<table>
<thead>
<tr>
<th>Language B.K.</th>
<th>American (L1) CS Humanities</th>
<th>Dutch (L2) CS Humanities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number of Sentences With ST Syntax</td>
<td>6.7 5.9</td>
<td>7.5 6.4</td>
</tr>
</tbody>
</table>

Note: Maximum Score = 10  
Above 5.0 = preference for ST syntax  
B.K. = Background Knowledge
The ANOVA analysis showed a significant overall effect due to the variance between the cells: $F(3,92) = 5.9, p < .005$. The Mann-Whitney U procedure demonstrated significant differences between American CS and humanities majors ($U = 2.0, p < .05$) and between Dutch CS and humanities majors ($U = 2.9, p < .005$) and no significant differences between American and Dutch CS students and between American and Dutch humanities students, although the Dutch used slightly more ST syntactic structures than the Americans, with a highly significant difference ($U = 4.2, p < .005$) between the Dutch CS students and the American humanities students. The Dutch CS majors used the most ST syntactic structures with an average of 7.5 out of 10. Both L1 and L2 Computer Science students opted for significantly more ST syntactic structures than did humanities students. A native knowledge of English produced more CL syntactic structures although not at a significant rate.

2.4 FIT/TUE 2--Writing Within an ST and CL Text (CL Cotext)

In the first study, while it was possible to make either CL or ST syntactic choices in each sentence, both L1 and L2 experts and nonexperts in computer science gave priority to the ST over the CL register in syntactically structuring their sentences, suggesting that the ST register is a function of the text topic of the material in addition to background knowledge. However, since the subjects were only working within ST material, an important question was raised as to the influence of the surrounding cotext on the syntactic choices of the writers. It could have been that the linguistic cotext of that material might indicate a priming effect (See Estival, 1985) towards the ST register produced by reading the surrounding text.

The second study, therefore, attempted to refine the empirical process to answer the following research question: Do syntactic structures chosen by L1 and L2 writers depend on a cognitive factor within the writer,
i.e., the background knowledge of the topic of the text--herein scientific vs. nonscientific--or by a priming effect from that subject's linguistic cotext? If the ST register is a function of ST knowledge, experts would use ST syntax and nonexperts would use CL syntax in an ST text, whereas both would use CL syntax in a CL text. If the ST register is a contextual function, both specialists and nonspecialists would produce ST syntax in ST text and CL syntax in CL text. Thus, in this study, an attempt was made to isolate the effect of the cotext (ST or CL) from the cognitive factor in the writer's decision to use CL or ST syntax in both native (L1) and nonnative (L2) settings. (For a full description of this study, see Strother & Ulijn, 1990.)

To balance the possible priming effect from the ST syntax in the computer science journal article, an additional contrasting text was chosen for this study, a work of fiction, which was already written in CL syntax. To eliminate a possible learning effect from the ST linguistic cotext in the computer science article (which is a natural consequence of the ST semantic context), the computer science article was rewritten into CL syntax.

The subjects were 48 American and 48 Dutch students--half with computer science background and half without. Each subject had to complete ten sentences which occurred within the two texts, both written in common language, but one having a scientific and one having a nonscientific topic (computer science vs. fiction).

The two independent variables are the subjects' language knowledge (L1 and L2) and background knowledge of the topic (CS vs. humanities). The dependent variable is the number, ranging from 0 to 5, of syntactic choices made by each group of subjects in each text (ST and CL). The nature of this data, the same writers working on two different texts, called for a simple \( t \) test analysis for all relevant factors and conditions.

As can be seen from Figure 2.4, both American (L1) and Dutch
Figure 2.4 Syntax Use in Computer Science and Fiction Texts for FIT/TUE 2

Key

<table>
<thead>
<tr>
<th>Context</th>
<th>Background Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Computer Science (ST)</td>
<td>Computer Science</td>
</tr>
<tr>
<td>○ Fiction (CL)</td>
<td>Humanities</td>
</tr>
</tbody>
</table>

Note: Above 2.5 = preference for ST syntax
      Below 2.5 = preference for CL syntax
(L2) subjects--CS and humanities--preferred ST syntax in the computer science (ST) text and CL syntax in the fiction (CL) text (p < 0.001). In the data, a number above 2.5 shows a preference for ST syntax and below 2.5, a preference for CL syntax.

Some interesting observations can be made, especially within each language group. For the American subjects writing within a scientific topic, background knowledge had very little effect. When writing within the computer science text, both CS and humanities subjects chose the ST syntax with almost the same frequency (3.7 vs. 3.6); therefore, the difference was not significant. The same was true for choice of CL syntax in the fiction context (1.7 for each group). However, for the Dutch subjects, the differences within each text were more notable. There is certainly no exclusive use of ST syntax by the Dutch. They have mastered both the ST and CL registers and use them in the appropriate semantic context derived from the topic of the text--fiction or computer science. However, in contrast to the Americans, in the CS text, Dutch CS subjects use more ST syntax than Dutch humanities subjects do (3.8 vs. 3.2) (p < 0.05). This result is plausible since, in a CS text, Dutch CS students profit largely from their prior exposure to English CS texts through their education. In this study, however, there is definitely no priming effect from the linguistic cotext since the CS text was written using CL syntax.

2.5 FIT/TUE 3--Writing Within a CS and Fiction Text (CL Cotext or ST Cotext)

The third study attempted to further refine the empirical process to more definitively answer the same basic research question: Do syntactic structures chosen by L1 and L2 writers depend on a cognitive factor within the writer, i.e., the background knowledge of the text's topic--herein scientific vs. nonscientific--or by a priming effect from that subject's linguistic cotext? If all subgroups were to choose predominantly ST syntax
in the CS text and CL syntax in the fiction text, it would provide a strong confirmation that the topic of the text determines choice of syntax.

The hypothesis was that all subjects--both L1 and L2, expert and nonexpert--would use significantly more ST syntactic structures in a scientific text than in a nonscientific text, therefore, that the topic of the text rather than linguistic cotext (syntactic structure) has a significant influence on the syntax chosen by the writer.

The L1 subjects of this study were 48 American students at Florida Institute of Technology--half with computer science background and half without. The L2 subjects were 24 computer science students at Eindhoven University of Technology and 24 humanities (psychology) students at Brabant Catholic University in Tilburg.

One of the goals of this study was to further remove the influence of the surrounding cotext from each writer's decision to choose particular syntactic structures to complete the sentences.

For this experiment, each subject had two texts to work with, the computer science journal article and the fictional short story. For the study, the computer science text was rewritten using mainly CL syntax and the fiction was rewritten (with sincere apologies to Jack London) using mainly ST syntax. Half of the subjects had both texts written in the ST register and half had both texts written in the CL register. As in the previous tests, within each text, five sentences were selected for the subjects to complete. In each of these sentences, it was possible to make either CL or ST syntactic choices. Subjects were given the words needed to complete the sentences in alphabetical order (violated syntax) and were given the same set of directions as in the previous studies. (See Appendix A for the complete test.)

A 2 x 2 x 2 x 2 ANOVA was used with the following independent variables:
The dependent variable was the number, ranging from 1 to 5, of ST syntactic choices made by each group of subjects in each text (ST and CL).

As can be seen from Tables 2.3 and 2.4, both American (L1) and Dutch (L2) subjects—both with expert and nonexpert backgrounds—preferred ST syntax in the computer science text and CL syntax in the fiction text (p < .001 for the Dutch and p < .0001 for the Americans), thereby supporting the hypothesis that the semantic context of a sentence (related to the topic), rather than background knowledge of the topic of the text itself, is the main factor which determines the syntax chosen by the writer. For both the American and the Dutch subjects, neither background knowledge of the topic being read (computer science) nor the linguistic context of the material had a significant effect on the kind of syntax used.

In either of the texts, an average score of 2.5 indicates a preference for ST syntax; a score below 2.5 indicates a preference for CL syntax. Some interesting observations can be made, especially within each language group. Figure 2.5 clearly demonstrates that, for all groups of students, the use of ST syntactic structures is significantly higher in the computer science context. It is interesting to note that, while the Dutch did significantly differentiate syntactic choices between the fiction and CS texts (2.2 vs. 3.7), they used more ST syntax even within the fiction text than did...
Table 2.3

Results for Syntax Use by American (L1) Subjects for FIT/TUE 3

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variables</th>
<th>Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ST syntax used</td>
<td>Comprehension</td>
</tr>
<tr>
<td>Background Knowledge</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>CS vs. Humanities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text Topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS vs. Fiction</td>
<td>( p &lt; .0001^{**} )</td>
<td>N.S.</td>
</tr>
<tr>
<td>Linguistic Context (Register)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST vs. CL</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

** Highly Significant
N.S. = Not Significant
## Table 2.4

**Results for Syntax Use by Dutch (L2) Subjects for FIT/TUE 3**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>ST Syntax Used</th>
<th>Comprehension</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>Text Topic</td>
<td>p &lt; 0.001 **</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>Linguistic Cotext</td>
<td>N.S.</td>
<td>p &lt; 0.01 *</td>
<td>p &lt; 0.01 *</td>
<td></td>
</tr>
<tr>
<td>Linguistic Cotext x Background Knowledge (Interaction)</td>
<td>p &lt; 0.025 *</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>Linguistic Cotext x Text Topic (Interaction)</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>Language Knowledge</td>
<td>p &lt; 0.001 **</td>
<td>p &lt; 0.001 **</td>
<td>p &lt; 0.001 **</td>
<td></td>
</tr>
<tr>
<td>Linguistic Cotext x Language Knowledge (Interaction)</td>
<td>p &lt; 0.001 **</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
</tbody>
</table>

* Significant  
** Highly Significant  
N.S. = Not Significant
Figure 2.5 Syntax Use in CS vs. Fiction Texts by L1 and L2 Subjects for FIT/TUE 3

- O - American
- □ - Dutch

Note: Above 2.5 = preference for ST syntax
Below 2.5 = preference for CL syntax

42
the American subjects (2.2 vs. 0.8 out of a possible 5). Therefore, for the Dutch subjects, the effect of language background caused them to use significantly (p < .001) more ST syntax than did the American subjects; the American subjects used an overall average of 4.3 while the Dutch used an average of 5.8 (out of a possible 10) ST syntactic structures.

When examining the effect of language knowledge and linguistic cotext (ST vs. CL) on syntactic choices, the picture is somewhat different for the Dutch and American subjects. Overall, the Dutch used more ST structures than did the American subjects. The Dutch used 5.5 ST structures in the CL context and 6.1 in the ST context. The Americans, however, used fewer ST structures overall--3.8 in the CL context and 4.7 (out of a possible 10) in the ST context. Thus, while the effect of the topic of the text is a much stronger influence than the cotext, it appears the Dutch may experience a slight priming effect from the syntax of the surrounding cotext when making their choice of syntactic structures.

As can be seen in Figure 2.6, Dutch computer science students use predominantly ST syntax in both ST and CL contexts; however, Dutch humanities students use more ST syntax in a ST context and more CL syntax in a CL context. An interesting observation is that the Dutch humanities subjects use slightly more ST syntax in an ST context than do Dutch computer science subjects; there is a slight reverse effect in the CL context where computer science students use more ST syntax than humanities students.

It is clear that the Dutch have mastered both the ST and CL registers and use them in the appropriate semantic context derived from the topic of the text--fiction or computer science. However, it appears that, unlike the American subjects, they can be influenced by the syntax of surrounding text. This may perhaps be because the Dutch writers attempted to find stylistic support in the surrounding text because of their limited language knowledge, whereas the Americans choose their
Figure 2.6 Effect of Language Background, Linguistic Context, and Background Knowledge on Syntax Use for FTI/TUE 3
syntactic structures from the wide variety in their repertoire. The higher use of ST syntactic structures probably also indicates a learning effect from frequent prior exposure to English texts throughout their education, especially in scientific fields where ST syntax is more frequent.

2.6 Writing Paragraphs Within a Scientific and a Nonscientific Topic with No Cotext

In the three FIT/TUE studies referred to above, the subjects always wrote sentences within a given text. Since the influence of the cotext on the syntactic structures chosen by the writer was a possible contaminating factor of the previous three writing studies, the fourth empirical study was designed to completely eliminate the influence of cotext while attempting to determine the factors which cause a writer to choose certain syntactic structures within specific text topics.

The 110 subjects of this study were students at Florida Institute of Technology. There were 51 American (L1) subjects. The 59 mixed ESL subjects (L2) included the following language groups: Arabic (31%), Chinese (12%), French, Spanish, Thai (10% each), and German, Greek, Gujarati, Icelandic, Indonesian, Korean, Japanese, Portuguese, Telugu, and Turkish. Approximately half of the L1 and half of the L2 group had a strong scientific background and half did not. A subject was defined as an expert, having a strong scientific background, if that subject was either in the third or fourth year or was a graduate student at university in a science or engineering major. A subject was defined as a nonexpert, having a nonscientific background, if he or she were in the first year of university and/or were majoring in a nonscientific field such as humanities or management. Note that the selection criteria for this study were broader than for the previous writing studies. In FIT/TUE 1-3, experts had a strong background in a specific field, computer science. In this study, however, the general areas of science and engineering were used for experts. See
Table 2.5 for a description of subjects of the study.

Power analysis was used to determine the number of subjects required for the study. For $\alpha = .05$, a desired power of .80, and expected $R^2$ of .40, the formulas in Cohen and Cohen (1983, p. 118) yielded a requirement of about 100 subjects in order to detect an effect size of 5 percent for any partial correlation. The accessible population made it possible to test 122 subjects which was more than adequate for the desired power. In contrast to FIT/TUE 1-3, for this study an uneven number of subjects were used because of availability constraints within each category. The fact that multiple regression was used to analyze the data allowed for unequal subject cell sizes.

Each subject wrote two paragraphs of five sentences each, one with an ST topic and one without, using only a word list as a cue.
Therefore, each subject had complete freedom of choice of syntax and register with no surrounding text to provide any kind of influence, thereby completely eliminating the effect of cotext on choice of register.

Each subject was instructed to write a logical, well-formed paragraph containing five sentences on a particular topic using a word list as a cue for that topic. They were instructed that they could change the form of any of the given words but that all of the words on the given list had to be used. To attempt to eliminate undue influence on the verb form chosen by the writers, all words in each list were nouns and were listed in alphabetical order. (See Appendix B for the test booklets).

Each subject was to write one paragraph on the topic of "Gravity" and one on the topic of "Newspapers." Half of the subjects were given the "Gravity" paragraph as the first task and the other half had the "Newspaper" paragraph as their first task.

The word list for the "Gravity" paragraph was:
- earth
- Isaac Newton
- falling objects
- tide
- force of attraction
- theory
- moon

The word list for the "Newspaper" paragraph was:
- advertisements
- information
- business
- news of the day
- effect
- reporters
- feature

The sentences were analyzed by an impartial grader as follows:

If the main structure of the sentence contains either a passive, a nominalization, an infinitive, or a participle, mark it $S$.

If it does not, mark it $C$.

The hypothesis that subjects would use significantly more ST syntactic structures within a scientific topic than within a nonscientific topic
was tested for each of three variables:

1. Language Knowledge
   - L1 - English as a first language
   - L2 - English as a second language

2. Background Knowledge
   - Expert - Scientific background
   - Nonexpert - Nonscientific background

3. Topic of the text being written about
   - Scientific
   - Nonscientific

The results of the tests for the three variables were as follows. First, a paired t-test was performed to discover whether significant differences exist between the use of ST and CL syntactic structures in the two paragraphs. The results showed a highly significant difference \( p < .005 \) so the hypothesis that subjects would use significantly more ST syntactic structures within an ST context than within a CL context was supported.

Regression analysis was then used to determine whether a subject's having a scientific or nonscientific background had a significant impact on the kinds of syntactic structures used. Background knowledge (scientific vs. nonscientific) was not a significant factor in choice of syntax. However, when the regression was performed to determine the effect of language background on choice of syntax, the differences were highly significant \( p < .0001 \).

Table 2.6 and Figure 2.7 show the mean number of sentences containing ST syntactic structures used within each paragraph by each group of subjects. These results, analyzed by regression analysis, show that both L1 and L2 writers prefer ST syntactic structures within the scientific topic and use fewer ST structures within a topic which is nonscientific. The results also show that L1 writers used ST syntactic structures more frequently than did L2 writers in both paragraphs, but especially within the paragraph having the scientific topic.
Figure 2.6 Effect of Language Background, Linguistic Context, and Background Knowledge on Syntax Use for FIT/TUE 3

- American CS
- American Humanities
- Dutch CS
- Dutch Humanities
Figure 2.7 Interaction of Language Background and Background Knowledge on Syntax Use in Scientific and Non-Scientific Paragraphs

- L1 Experts
- L1 Nonexperts
- L2 Experts
- L2 Nonexperts

Note: Above 2.5 = preference for ST syntax
Below 2.5 = preference for CL syntax
2.7 Comments on the Procedures

One of the main limitations of each of the FIT/TUE studies is the amount of and way in which the sentences in the cotext have been rewritten since that, of course, influences the results. To make the study valid, a strong distinction had to be made between ST and CL syntax in each of the texts. It is, of course, ideal to use authentic text whenever possible. In studies such as these, when rewriting text, one strives to avoid having authentic text become artificial from overediting. Thus, a continuum develops between authentic or "natural" text and completely rewritten or artificial text. In these studies, if every sentence had been rewritten and thus the distinction between ST and CL cotext became even stronger, perhaps the results would have been biased because of the artificial syntactic frequency loads. In addition, the choice of sections of each sentence to be presented in violated syntax was a decision which had to be made. Because of the desire to further test the reading-writing connection, one of the criteria for selecting sentences for rewriting was the possibility of using the information from those sentences to test reading comprehension. (See Chapter III for a complete discussion of this procedure.)

Thus, in these FIT/TUE studies, there was a constrained cotext in which subjects wrote, and all writing was done at the sentence level. While this put a strong limitation on a writer's syntactic choices, this constraint made it easier to evaluate and quantify results since all subjects worked within the same limits. However, even though the subjects wrote at the sentence level, they were writing within a complete text and thus were analyzing the text beyond the sentence level to complete their tasks.

In the study in which subjects wrote scientific and nonscientific paragraphs, the question of cotext was completely eliminated and thus subjects were completely free to make whatever syntactic choices they wished within each topic. As a result, there was a wide variety of sentence
length and sentence complexity among the subjects. This made evaluation and quantification of the syntactic structures quite difficult and the subjects' paragraphs were not perfectly comparable. However, the independent rater made every effort to judge the syntactic choices of the writers as consistently as possible.

The final paragraph writing study must be compared with the FIT/TUE studies with a degree of caution because of the differences in (1) the structure of the tests themselves—subjects' working with violated syntax within a context vs. writing paragraphs with no context; (2) L2 subjects' first languages (mixed ESL); and (3) background knowledge of the subjects—computer science vs. mixed science and engineering; (4) the style of the CL text used (expository in the paragraph on the newspaper topic vs. narrative fiction in FIT/TUE 2 and 3).
CHAPTER III
THE EFFECT OF ST VS. CL SYNTAX ON READING

3.1 Psycholinguistics and Reading Comprehension

The critical connection between the writer and the reader is the text. Chapter II discussed the writer's choice of syntax when writing a text on a particular topic. This chapter examines the effect of that syntactic choice on the reader. This book takes a psycholinguistic approach to reading comprehension in an attempt to examine how a reader processes a text and to examine the reader's cognitive processes as he or she attempts to understand the content of a text. By focusing on syntax, the study seeks to determine what role that particular facet of text plays in the comprehension process.

Some characteristics of human information processing, for example, reading, are very pervasive. Among these are the concepts of hierarchical organization, bottom up versus top down processing, and parallel processing. Hierarchical organization is an essential feature of human language. When a person reads, rather than simply identifying individual words, he or she integrates words into higher order linguistic units of syntactic categories. Individual words are organized into higher order categories or constituents, such as nouns, verbs, or prepositional phrases. These are then organized into sentences which may be organized into a larger linguistic unit such as a story (Glass & Holyoak, 1986).

Evidence of hierarchical feature analysis during reading was obtained by Healy (1981) in her work with college students. Perception of words may be at the level of features, letters, or the entire word; thus, the visual processing system is very flexible. This flexibility in reading results from the interaction of two kinds of analytical processes existing at different
levels in the hierarchy, with sensory analysis of the input at the bottom of the hierarchy and abstract semantic representations at the top. The resulting kinds of processing are called bottom up and top down. The defining property of a strictly bottom up process is that the outcome of a lower step is never affected by a higher step in the process. On the other hand, if the output of a lower step is influenced by a higher one, it is called top down processing (Glass & Holyoak, 1986).

In the reading process, several different procedures have to occur at once; these include eye movement, visual identification, and sentence construction. These different procedures are coordinated so they all occur at the same time but for different parts of the input. Thus reading involves successive recoding of the input through hierarchically organized levels of representation involving both bottom up procedures, based directly on the sensory input, as well as top down procedures which use higher levels, such as expectations generated by the context. In addition to bottom up and/or top down, a great deal of parallel or simultaneous processing occurs during the reading process. Parallel processing is defined as a processing procedure that can perform more than one act on an input at a time. (Glass & Holyoak, 1986).

There is a constant interaction between form and meaning in both reading and writing. Figure 3.1 shows a comparison between serial, parallel, and partial parallel processing for both writing and reading.

The psycholinguistic model of reading used in the studies in this book is Ulijn's (1987) partial parallel reading comprehension model. In this model, shown in Figure 3.2, rather than a strictly top down or bottom up mode of processing, there is interaction among the components of the cognitive systems and the knowledge sources. The studies reported herein support a version of that model in which only a superficial syntactic analysis occurs first and a more thorough syntactic analysis is undertaken only if the interaction between conceptual and lexical analysis fails to
The Operation of Text & Sentence Parser in Reading

The Type of Text & Sentence Generator in Writing

1. Serial

- Top Down

   \[
   \text{Text} \rightarrow \text{SA} \rightarrow \text{LA} \rightarrow \text{CA} \rightarrow \text{Text}
   \]

   or

   \[
   \text{CA} \rightarrow \text{LA} \rightarrow \text{SA} \rightarrow \text{Text}
   \]

2. Parallel

   \[
   \begin{align*}
   \text{CA} \rightarrow \text{LA} \rightarrow \text{SA} \\
   \text{LA} \rightarrow \text{Text} \\
   \text{SA} \rightarrow \text{Text}
   \end{align*}
   \]

   Interactive

3. Partial Parallel

   \[
   \begin{align*}
   \text{Text} \rightarrow \text{SA} \\
   \text{LA} \rightarrow \text{SA} \\
   \text{CA} \rightarrow \text{Thorough}
   \end{align*}
   \]

   Interactive

   \[
   \begin{align*}
   \text{C} \rightarrow \text{L} \rightarrow \text{S} \\
   \text{C} \rightarrow \text{Text}
   \end{align*}
   \]

   Revision

\textbf{Figure 3.1 Cognitive Processing Models for Reading and Writing}

\textbf{Legend:}

- \( C \) = conceptual retrieval
- \( CA \) = conceptual analysis
- \( L \) = lexicalization
- \( LA \) = lexical analysis
- \( S \) = syntactic structuring
- \( SA \) = syntactic analysis

\textbf{Source:} Ulijn 1987

55
Figure 3.2: A Cognitive Model of the Reader
reach comprehension of the text.

This raises some of the key questions of this book: what is the influence of syntax, particularly syntactic structures common to the ST register, on readability? At what point is only a superficial syntactic analysis adequate for comprehension? At what level of L2 language knowledge does syntax play a role in readability?

At least a basic knowledge of syntax is required for reading comprehension. According to Nilagupta (1977, p. 585), "The rules of syntax are not just the rules that the writer applies to organize his statements-- they are the rules he assumes the receiver knows in order to be able to extract the meaning from statements."

Alderson and Alvarez (1978, p. 4) stated "It is clear that grammatical information can convey meaning. Identifying words as nouns, not adverbs, presumably helps one to divide up the sentence and get at the syntax, which helps one to make predictions as to what the sentence might mean." However, the amount of analysis of syntax done by the reader in certain situations may vary considerably, according to such psycholinguists as Hatch (1983) and Ulijn (1984). Both have analyzed linguistic factors which are specific to second language reading comprehension strategies using psycholinguistic models of reading. These models "clearly distinguish several linguistic levels operating in reading: textual, syntactic, morphemic, lexical, graphophonemic or graphemic, which have been experimentally evidenced in reading" (Ulijn, 1980, p. 453). Even though there is interaction among these linguistic factors, there is the need to isolate the individual factors which contribute to problems with reading comprehension.

This chapter summarizes several important studies which have been conducted to examine the effect of syntax on readability. Three of the reading tests were conducted as part of the writing experiments which are reported in full in Chapter II. (Ulijn & Strother, 1987; Strother & Ulijn,
It is important to note that these tests were designed as writing experiments; however, some aspects of the studies were used for an analysis of reading comprehension. Briefly, the subjects of these three writing experiments had to complete sentences within a text using words or groups of words which were presented in alphabetical order (violated syntax). While they were attempting to comprehend the entire text, they had to read these same sentences with violated syntax; therefore, they had to base their reading comprehension of these particular sentences on figuring out the meaning with incomplete information. So, it was a problem-solving task in addition to being a reading comprehension task.

The other two were strictly reading tests which contrasted reading comprehension of ST vs. CL syntax. In the first of these studies (Ulijn & Strother, 1985; Strother & Ulijn, 1987; Ulijn & Strother, 1990), all groups of subjects read the computer science article in either ST or CL syntax and in the second study, all groups read the fiction text in either ST or CL syntax.

To allow comparison among the studies, all experiments used the same ST text, a slightly shortened version of a computer science article that had no mathematical formulas or illustrations. The article, "Mass Storage Systems and Evolution of Data Center Architecture," (Miller, 1982) came from the periodical *Computer*. Where applicable, the same CL text, a shortened version of Jack London's fictional short story, "To Build a Fire" (1964), was used.

Again, for comparison purposes, the same general categories of subjects were used for all of the reading and writing studies. In each study, both native speakers of English (L1) and nonnative speakers of English (L2) were used. In each study, half of each L1 or L2 group had expertise in the computer science field or had a strong scientific background. The other half of each group had neither a background in computer science nor a strong scientific background. The precise operational criteria for the
expert/nonexpert category varied somewhat depending on the study. In addition, in some studies, the L2 subjects were either Dutch or mixed ESL (English as a Second Language) or both. The specific characteristics of each subject group are defined for each of the major studies.

As was discussed in the literature review in Chapter I, there are a number of factors which affect readability. As an overview then, these studies examine the effects of language (L1 vs. L2), background knowledge of the text (expert vs. nonexpert), and syntactic structure of the text (ST vs. CL) on the readability of a certain text. For the purpose of these studies, four syntactic structures have been isolated—the passive, nominalizations, participles, and infinitives—because of their frequency in the ST register.

3.2 Reading Violated Syntax

As was discussed in Chapter II, the first three writing studies were cooperative efforts between Florida Institute of Technology (FIT) and Eindhoven University of Technology (TUE). Each of those writing studies included a reading component in which subjects were given words or groups of words, presented in alphabetical order (violated syntax), within a text. Thus, subjects in these three studies had to solve the problem of putting the sentence elements in order to create meaning and then read the results. The details of each FIT/TUE study are presented below.

3.2.1 FIT/TUE 1

Table 3.1 shows a distribution of the subjects for this study and the two which follow.

In the first study—FIT/TUE 1—(see Strother & Ulijn, 1989a, for a complete description of this study), all subjects read the computer science journal article which had 10 sentences within the text written with violated syntax. The subjects' comprehension of the material was tested by having
Table 3.1

Subjects of FIT/TUE Writing/Reading Studies

<table>
<thead>
<tr>
<th>Background Knowledge</th>
<th>FIT/TUE 1</th>
<th>FIT/TUE 2</th>
<th>FIT/TUE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Knowledge</td>
<td>L1 L2</td>
<td>L1 L2</td>
<td>L1 L2</td>
</tr>
<tr>
<td>Experts</td>
<td>24 24</td>
<td>24 24</td>
<td>24 24</td>
</tr>
<tr>
<td>Nonexperts</td>
<td>24 24</td>
<td>24 24</td>
<td>24 24</td>
</tr>
<tr>
<td>Total</td>
<td>48 48</td>
<td>48 48</td>
<td>48 48</td>
</tr>
<tr>
<td>n = 96</td>
<td>n = 96</td>
<td>n = 96</td>
<td></td>
</tr>
</tbody>
</table>

them answer 10 true/false questions based on the 10 sentences which had been written with violated syntax. The reading comprehension scores for this study are contrasted with those of similar subjects who read the same text without violated syntax (Strother & Ulijn, 1987, discussed below).

The data presented in Table 3.2 demonstrate to what extent sentences can be understood with and without violated syntactic structure. While all groups showed some differences between non-violated (both CL and ST) and violated syntax, the Dutch humanities students showed the greatest loss of reading comprehension with violated syntax. However,
### Table 3.2

Mean Reading Comprehension Scores of a Computer Science Text for FIT/TUE 1

<table>
<thead>
<tr>
<th>Language</th>
<th>Type of Text</th>
<th>American L1 Comp. Sci. n = 24</th>
<th>American L1 Humanities n = 24</th>
<th>Dutch L2 Comp. Sci. n = 24</th>
<th>Dutch L2 Humanities n = 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.K.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Text</td>
<td></td>
<td>Nonviolated CL</td>
<td>Syntax</td>
<td>Violated Syntax</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.6</td>
<td>8.3</td>
<td>8.1</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2</td>
<td>7.3</td>
<td>8.1</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.7</td>
<td>9.3</td>
<td>9.5</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Note: Maximum score = 10  
B.K. = Background Knowledge

### Table 3.3

Mean Times (in minutes) to Read the Text and Complete the Test for FIT/TUE 1

<table>
<thead>
<tr>
<th>Language</th>
<th>Type of Text</th>
<th>American L1 Comp. Sci. n = 24</th>
<th>American L1 Humanities n = 24</th>
<th>Dutch L2 Comp. Sci. n = 24</th>
<th>Dutch L2 Humanities n = 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.K.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Text</td>
<td></td>
<td>Nonviolated CL</td>
<td>Syntax</td>
<td>Violated Syntax</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.4</td>
<td>9.5</td>
<td>16.4</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.3</td>
<td>10.8</td>
<td>17.3</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.0</td>
<td>9.8</td>
<td>9.6</td>
<td>14.0</td>
</tr>
</tbody>
</table>

Note: Reading times relate to the whole text, not only to the 10 violated or nonviolated sentences.  
B.K. = Background Knowledge
comprehension of both violated and non-violated syntax is very high. It was found that there were no significant differences in comprehension among the groups, whether controlling for language background, background knowledge of the subject to be tested, or register of the text itself.

Reading times for the whole text, not just the selected test sentences, were also compared among the same groups (see Table 3.3), and while these times are not perfectly reliable because they were recorded by the subjects themselves, they are an indication of differences among the groups. Not surprisingly, L2 readers always took longer to read the text. A striking finding is that CS (computer science) students tended to take significantly more time than humanities students reading a CS text. This is especially true for the Dutch CS students reading violated syntax. Perhaps specialists in a field work harder at eliciting the real meaning from the text and can process it at a deeper level because they can tie it more quickly and easily into their background knowledge, whereas those who do not know the subject skim over the surface level meaning, therefore reading faster.

3.2.2 FIT/TUE 2

In FIT/TUE 1, the subjects read only a computer science journal article. In an attempt to make the results more generalizable, a second study, FIT/TUE 2 (see Strother & Ulijn, 1990, for a complete description of this study) was designed to have subjects read two authentic texts, one with a scientific topic (the computer science article) which was written in ST syntax, and one with a nonscientific topic (the fiction text) which was written in CL syntax. The design of the test was the same as that of FIT/TUE 1.

Subjects' comprehension of the material was tested by having them answer 10 true/false questions, 5 on each text, based on the sentences which were presented with violated syntax. Comprehension
scores of both the computer science and fiction texts were measured for each group (see Table 3.4). There is a significant overall effect of background knowledge on comprehension. The most notable exception is that the American expert (computer science) subjects had significantly higher comprehension scores than the American nonexpert (humanities) subjects on the fiction text (4.71 vs. 4.29). There is a possible ceiling effect, but it is rather surprising that the American computer science subjects understood the fiction text better than the humanities subjects. At least for this group of 48 readers, the CS subjects appear to be better overall readers than the humanities subjects.

There is a general effect of language knowledge on comprehension. The American subjects (ranging from 4.1 to 4.7) had higher comprehension scores than the Dutch subjects (ranging from 3.4 to 3.8) on both the CS and the fiction texts (p < 0.001, p < 0.002). This result is probably explained by the American subjects' command of English as a native language. In contrasting the American and Dutch humanities subjects, this difference is not significant in reading the CS text (4.1 vs. 3.8). This exception is not surprising; both have a lack of knowledge about the subject matter or text topic which makes comprehension of that text difficult, so language knowledge is not very helpful there.

Finally, there is no effect at all within groups of text topic (computer science vs. fiction) so both texts were equally difficult to understand for these groups of readers.

Reading times, again recorded by the subjects themselves, give plausible results, as shown in Table 3.5. The Dutch subjects, reading in English as a second language, took longer to read than did the Americans (L1) (p < 0.01 and p < 0.05). However, again looking at the Americans, the humanities subjects (nonexperts) took significantly longer (p < 0.02) to read both the CL and the ST text than did the CS (expert) subjects. In this case, this correlates with their lower reading scores, although length of
Table 3.4

**Mean Reading Comprehension Scores of Computer Science and Fiction Texts for FIT/TUE 2**

<table>
<thead>
<tr>
<th>Language B.K. Topic</th>
<th>American (L1)</th>
<th>Dutch (L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>Humanities</td>
<td>CS</td>
</tr>
<tr>
<td>Computer Science (ST)</td>
<td>4.5</td>
<td>3.7</td>
</tr>
<tr>
<td>Fiction (CL)</td>
<td>4.7</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Note: Maximum score = 5

Table 3.5

**Mean Times (in minutes) to Read the Text and Complete the Test for FIT/TUE 2**

<table>
<thead>
<tr>
<th>Language B.K. Time</th>
<th>American (L1) CS</th>
<th>Humanities</th>
<th>Dutch (L2) CS</th>
<th>Humanities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number of Minutes to Complete Text</td>
<td>25.7</td>
<td>30.5</td>
<td>33.0</td>
<td>36.1</td>
</tr>
</tbody>
</table>
time taken to read a text does not always correlate with reading comprehension of that text. While the Dutch humanities subjects also took longer than the CS subjects, the differences were nonsignificant. In this reading comprehension study, for most subjects, lower comprehension scores correlated with longer reading times, whereas in the previous one, for CS subjects and for Dutch subjects, higher comprehension scores correlated with longer reading times. As expected, nonnative readers took longer to process the text than native readers.

A summary of results for the three variables, background knowledge, language knowledge, and text topic, is given in Table 3.6.

### 3.2.3 FIT/TUE 3

The third study, FIT/TUE 3, also incorporating the need for subjects to read violated syntax, replicated the design of the previous reading studies as part of a more complex experiment (Strother & Ulijn, 1989b) in which each subject read both the computer science journal article and the fiction texts. The computer science text was rewritten using mainly CL syntax and the fiction was rewritten using mainly ST syntax. Half of the subjects had both texts written in the ST register and half had both texts written in the CL register. As in the previous tests, subjects had to answer 10 true/false questions, 5 on each text.

Comprehension scores of both the computer science and fiction texts were measured for each group. The American readers had significantly higher comprehension scores (8.4 out of 10) than did the Dutch readers (6.7 out of 10) (p < .001). This was true whether they were reading the computer science article or the fiction text, whether they were reading in the ST or CL register. Within the two language groups, some interesting differences emerge. For the Dutch, while neither the reader's background knowledge (CS vs. humanities) nor the topic of the text itself (CS vs. fiction) had a significant effect on reading comprehension scores,
### Table 3.6

Results of Reading Comprehension Scores for FIT/TUE 2

<table>
<thead>
<tr>
<th>Background Knowledge</th>
<th>ST Syntax Use</th>
<th>Comprehension</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>L₂CS/CL - L₂H/CL</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₁CS/CL - L₁H/CL</td>
<td>N.S.</td>
<td>p &lt; 0.05</td>
<td>N.S.</td>
</tr>
<tr>
<td>L₂CS/ST - L₂H/ST</td>
<td>p &lt; 0.05</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>L₁CS/ST - L₁H/ST</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₂CS - L₂H</td>
<td>N.S.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L₁CS - L₁H</td>
<td></td>
<td>p &lt; 0.05**</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language Knowledge</th>
<th>ST Syntax Use</th>
<th>Comprehension</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>L₂CS/CL - L₁CS/CL</td>
<td>N.S.</td>
<td>p &lt; 0.001**</td>
<td></td>
</tr>
<tr>
<td>L₂H/CL - L₁H/CL</td>
<td>N.S.</td>
<td>p &lt; 0.001**</td>
<td></td>
</tr>
<tr>
<td>L₂CS/ST - L₁CS/ST</td>
<td>N.S.</td>
<td>p &lt; 0.002**</td>
<td></td>
</tr>
<tr>
<td>L₂H/ST - L₁H/ST</td>
<td>N.S.</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₂CS - L₁CS</td>
<td></td>
<td>p &lt; 0.01*</td>
<td></td>
</tr>
<tr>
<td>L₂H -  L₁H</td>
<td></td>
<td>p &lt; 0.05*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semantic Context/Topic</th>
<th>ST Syntax Use</th>
<th>Comprehension</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>L₂CS/CL - L₂CS/ST</td>
<td>p &lt; 0.001**</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₂H/CL - L₂H/ST</td>
<td>p &lt; 0.001**</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₁CS/CL - L₁CS/ST</td>
<td>p &lt; 0.001**</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>L₁H/CL - L₁H/ST</td>
<td>p &lt; 0.001**</td>
<td>N.S.</td>
<td></td>
</tr>
</tbody>
</table>

**Key:**  
- **L₂CS** Dutch Computer Science students  
- **L₁CS** American Computer Science students  
- **L₂H** Dutch Humanities students  
- **L₁H** American Humanities students  
- **CL** Common language—Fiction context  
- **ST** Scientific/Technological language—Computer Science context

Significance levels:  
- * = Significant  
- ** = Highly Significant  
- N.S. = Not Significant
the linguistic cotext did have a significant effect ($p < .01$) as shown in Figure 3.3. The Dutch read the CL syntax with higher comprehension than they did the ST syntax. That in itself is surprising because, as was discussed in Chapter II, the Dutch preferred to write ST syntax in both text topics (See Table 3.7). For the American subjects, however, there was no significant effect of background knowledge, linguistic register of the text, or text topic on their reading comprehension scores (See Table 3.8).

Times to complete the test were also recorded, and while these times are not perfectly reliable since they were recorded by the subjects themselves, they are a strong indication of differences among the groups. It is not surprising that the Dutch, reading English as a Second Language, took significantly longer (an average of 37.8 minutes) to read than did the Americans (an average of 25.4 minutes) reading in their native language ($p < .001$). Figure 3.4 shows the effects of linguistic cotext (ST vs. CL) on reading times. For the American subjects, linguistic cotext did not have a significant influence; as a matter of fact, the times for the ST cotext (25.7) and for the CL cotext (25.6) were very close. However, the Dutch subjects spent significantly more time on the ST cotext than on the CL cotext for both the computer science and the fiction texts ($p < .01$). Thus it can be concluded that, for the Dutch, the ST cotext is more difficult; there is lower comprehension of the text and more time needed to complete the test.

This confirms earlier results (Strother & Ulijn, 1989a, Strother & Ulijn, 1990) that L1 reading is better and more efficient, showing a possible correlation between comprehension and time. An earlier finding (Ulijn & Strother, in press) that the Dutch understood better yet read more slowly was not supported in this work. However, in all three of the FIT/TUE writing studies, subjects read violated syntax which gave the Dutch more problems; the Americans, on the other hand, seemed to be able to overlook the violated syntax and get at the meaning successfully.
Figure 3.3 Effect of Linguistic Context on Reading Comprehension Scores for Fiction Texts for FIT/TUE 3

○ - American
■ - Dutch
### Table 3.7

Results for Reading Comprehension Scores by Dutch (L2) Subjects for FIT/TUE 3

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variables</th>
<th>ST syntax used</th>
<th>Comprehension</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>N.S</td>
<td>N.S</td>
<td>N.S</td>
<td></td>
</tr>
<tr>
<td>Text Topic</td>
<td>p &lt; 0.001 **</td>
<td>N.S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic Context</td>
<td>p &lt; 0.01 *</td>
<td>p &lt; 0.01 *</td>
<td>p &lt; 0.01 *</td>
<td></td>
</tr>
<tr>
<td>Linguistic Context x Background Knowledge (Interaction)</td>
<td>p &lt; 0.025 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic Context x Text Topic (Interaction)</td>
<td>N.S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Knowledge</td>
<td>p &lt; 0.001 **</td>
<td>p &lt; 0.001 **</td>
<td>p &lt; 0.001 **</td>
<td></td>
</tr>
<tr>
<td>Linguistic Context x Language Knowledge (Interaction)</td>
<td>p &lt; 0.001 **</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant  
** Highly Significant  
N.S. = Not Significant
Table 3.8

Results for Reading Comprehension Scores by American (L1) Subjects for FIT/TUE 3

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>ST Syntax Used</th>
<th>Dependent Variables</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background Knowledge</td>
<td>ST Syntax Used</td>
<td>Dependent Variables</td>
<td>Time</td>
</tr>
<tr>
<td>CS vs. Humanities</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>Text Topic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS vs. Fiction</td>
<td>p &lt; .0001 **</td>
<td>N.S.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Linguistic Cotext (Register)</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>ST vs. CL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Highly Significant  
N.S. = Not Significant  
N.A. = Not Applicable
Figure 3.4 Effect of Linguistic Context on Mean Times to Complete the Test for FIT/TUE 3

- American
- Dutch
3.3 Reading ST vs. CL Syntax Within a Computer Science Topic

In addition to the above three studies which incorporated a reading section as part of a writing test, each including some violated syntax, there are two main reading studies, using only normal (nonviolated) syntax, which are part of this book. The first (Strother & Ulijn, 1987, Ulijn & Strother, 1990) dealt with the reading comprehension of ST texts and the other (reported herein in full detail) with that of CL texts. Both studies had the same research design and asked the same basic research questions. The research questions for the first study were:

1. Is the supposed complexity of syntax in an authentic (ST) text more difficult for readers to comprehend than a syntactically rewritten (CL) version of the same text?
2. To what extent does background knowledge affect syntactic analysis in reading an ST text?
3. To what extent does language knowledge affect syntactic analysis in reading an ST text?

The first study focussed on technical English with the hypothesis that significant differences would be obtained between scores on a reading comprehension test by university students who read an authentic (unsimplified) computer science journal article and those who read a syntactically simplified version of the same article. To revise (and supposedly simplify) the text, ten one-sentence passages, equally distributed over the entire text, were chosen for rewriting. They contained structures that occur with greater frequency in the ST register and that are reported to cause some problems: namely, passives, nominalizations, participles, and infinitives. These sentences were rewritten into a syntactic form more suitable for a common language version.
Care was taken not to change the meaning of any sentence. No lexical items or other elements of the original sentences were changed. Therefore, the same content load was maintained. To strengthen the internal validity of the tests, both the revised version of the article and the question set were checked by a computer science professional to ensure that changing the syntax in the text indeed did not change the meaning of the text and that the questions accurately queried information from the article.

The following two sentences from the article are examples of the syntactic simplification from the ST register to a common language register done for this research project:

1. **Authentic text:** Replacement of current devices with optical data disks at a higher system level is required; this implies considerable software development to take advantage of these promising device characteristics. **Revised text:** The user must replace current devices with optical data disks at a high system level. The programmer must develop considerably more software to take advantage of these promising device characteristics.

2. **Authentic text:** When such movement is possible under system control, the problem becomes one of separating and clustering sets of files independently of the physical volumes and providing a mechanism to have desired files more readily available. **Revised text:** When such movement is possible under system control, the problem is: can the system separate and cluster sets of files independently of the physical volumes and provide a mechanism to have desired files more readily available.
Two types of test booklets were made, one using the authentic article and the other using the partly rewritten version. Both were followed by the same true/false statements referring in random order to the 10 passages that had been rewritten in the text. (See Appendix C for the complete tests.)

The subjects of the study were from Florida Institute of Technology in Melbourne, Florida, and Eindhoven University of Technology in The Netherlands. The 48 American subjects had English as a native language (L1). The 115 English as a second language subjects (L2) included the following language groups: Dutch (21%), Chinese (19%), Arabic (17%), Spanish (13%), and other mixed ESL (30%). Within each group, approximately half were experts and half were nonexperts in computer science.

A 2 x 2 x 2 ANOVA (one way, with repeated measures) was used with the following independent variables (factors) and conditions:

1. Language knowledge
   - LI--English as a first language
   - L2--English as a second language

2. Background knowledge of the topic being read
   - Expert--computer science background
   - Nonexpert--No computer science background

3. Syntactic structure of the text being read
   - ST--Science and Technology
   - CL--Common Language

The dependent variables were the number of correct answers on the comprehension test and text reading time.

As can be seen in Table 3.9, language register of the text did not have a significant effect on reading comprehension scores, but both background knowledge and language knowledge did (p < .01 vs. p < .005). Thus, it can be concluded that the syntax of the text did not cause
### Table 3.9

**ANOVA Results for Subjects Reading a Computer Science Text**

<table>
<thead>
<tr>
<th>Main Effects of:</th>
<th>on Comprehension Scores</th>
<th>on Reading Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR: Language Register of the Text: Authentic vs. Syntactically Adapted</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>BK: Background Knowledge: Experts vs. Nonexperts</td>
<td>p &lt; .01*</td>
<td>N.S.</td>
</tr>
<tr>
<td>LK: Language Knowledge: Native vs. Nonnative English</td>
<td>p &lt; .005**</td>
<td>p &lt; .005**</td>
</tr>
</tbody>
</table>

**Interactions**

| LK x BK | N.S. | N.S. |
| LK x LR | N.S. | p < .005** |
| LK x BK x LR | N.S. | p < .005** |

* Significant  
** Highly Significant  
N.S. = Not Significant
significant differences in comprehension between the ST and CL versions, but that the domain relevant knowledge of computer science did help the subjects comprehend the computer science text significantly better. It can also be seen that language knowledge has a very significant effect on reading time. This factor is involved in three highly significant interactions: (1) with background knowledge ($p < .005$), (2) with language register of the text ($p < .005$), and (3) with background knowledge and language register ($p < .005$). These interactions are due to the fact that Americans read nearly twice as quickly as the Dutch.

Table 3.10 summarizes the mean number of correct answers for both L1 and L2 subjects and Table 3.11 summarizes the mean time in minutes to complete the task of reading the text and answering the questions.

In summary, the statistical analysis showed that, whether controlling for language background or background knowledge of the subject to be tested, the register of the text itself had no effect on reading comprehension scores. These results indicate that readers overlook all kinds of syntactic variants that have usually been assumed to be simplifications: nominalizations versus verb phrases, passive versus active constructions, and participle constructions versus subordinate clauses. In the experiment, the revisions from a scientific and technical register syntax into a more simplified common language syntax affected neither the comprehension scores nor the reading time. In addition, neither the background knowledge of computer science (experts vs. nonexperts) nor the specific language background of nonnative readers affected the scores or the reading time. Thus, the syntactic adaptation of the English computer science text did not really help subjects to comprehend better or to read more quickly, either for the natives (Americans) or for the nonnatives (Dutch, Chinese, Spanish, Arabic, or other language backgrounds).
Table 3.10

Mean Number of Correct Answers for L1 and L2 Subjects Reading a Computer Science Text

<table>
<thead>
<tr>
<th>Background Knowledge of Students</th>
<th>Language Register of Text</th>
<th>L1 Subjects</th>
<th>L2 Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST</td>
<td>CL</td>
</tr>
<tr>
<td>Experts in Computer Science</td>
<td>American English</td>
<td>8.3</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>9.3</td>
<td>9.7</td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
<td>8.9</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>n = 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonexperts in Computer Science</td>
<td>American English</td>
<td>7.3</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td>7.1</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td>n = 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.11
Mean Times (in minutes) for L1 and L2 Subjects
To Read a Computer Science Text

<table>
<thead>
<tr>
<th>Background Knowledge of Students</th>
<th>Language Background</th>
<th>L1 Subjects</th>
<th>L2 Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ST</td>
<td>CL</td>
</tr>
<tr>
<td>Experts in Computer Science</td>
<td>American English</td>
<td>9.5</td>
<td>11.4</td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonexperts Computer Science</td>
<td>American English</td>
<td>10.8</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>n = 24</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4 Reading ST vs. CL Syntax Within a Nonscientific (Fiction) Topic

This reading study, using a shortened version of Jack London's short story, "To Build a Fire" (1964), asked the same basic research questions as did the reading study using the computer science journal article; namely:

1. Is the syntax in an authentic CL text less difficult for readers to comprehend than a syntactically altered ST version of the same text?
2. To what extent does background knowledge affect syntactic analysis in reading an ST text?
3. To what extent does language knowledge affect syntactic analysis in reading an ST text?

The hypothesis was that significant differences would be obtained between scores on a reading comprehension test by university students who read an authentic common language fiction text and those who read a supposedly more difficult version of the same text, one with a high percentage of ST syntactic structures.

To rewrite the text, twenty-one sentences, equally distributed over the entire text, were rewritten into the ST register, using specifically 20 additional passives, 4 additional nominalizations, 2 additional participles, and 7 additional infinitives. Care was taken not to change the meaning of any sentence. No lexical or other elements of the original sentences were changed. Therefore, the same content load was maintained. To strengthen the internal validity of the test, both the revised version of the fiction text and the question set were carefully checked by a humanities professor to ensure that changing the syntax in the text indeed did not change the meaning of the text and that the questions accurately queried the information from the story.

The following sentences from the fiction text are examples of the syntactic simplification from the CL register into the ST register done for
this research project:

1. **Authentic text:** It was a steep bank, and he paused for breath at the top, excusing the act to himself by looking at his watch.
   **Revised text:** The steepness of the bank caused him to pause for breath at the top, excusing the act to himself by looking at his watch.

2. **Authentic text:** He was a warm-whiskered man, but the hair on his face did not protect the high cheekbones and the eager nose that thrust itself aggressively into the frosty air.
   **Revised text:** He was a warm-whiskered man, but his high cheekbones and the eager nose that thrust itself aggressively into the frosty air were not protected by the hair on his face.

Two types of test booklets were made, one using the authentic story (in CL syntax) and the other using the version which had been partly rewritten into ST syntax. Both were followed by the same true/false statements referring in random order to some of the passages that had been rewritten in the text. (See Appendix D for the complete text.)

As was discussed in Chapter II, power analysis was used to determine the number of subjects required for the study and multiple regression was used to analyze the data.

The 122 subjects of the experiment were students at Florida Institute of Technology in Melbourne, Florida. One group of students was composed of native speakers of English (L1) and the other group was composed of students with English as a Second Language (L2) with mixed first language backgrounds. The 59 L2 subjects included the following
language groups: Arabic (31%), Chinese (12%), French, Spanish, and Thai (10% each), Japanese (5%), German, Gujarati, Indonesian, Telugu, Turkish (3% each), and Greek, Icelandic, and Korean (2% each).

Since this test used a fiction text, the issue of background knowledge is different from the previously discussed studies and the categories "expert" and "nonexpert" in science or computer science are not really relevant to the fiction text. However, to allow comparison with all of the previously reported studies, these subject divisions were kept. Thus, approximately half of each group, having a strong scientific background, were labeled experts and the other half, having little or no scientific background, were labeled nonexperts. A subject was considered to have a strong scientific background if that subject was either in the 3rd or 4th year or was a graduate student in a science or engineering major at the university. A subject was considered to have little or no scientific background if he or she was either in the first year of university or was majoring in a nonscientific field such as humanities or management. The L2 group had an advanced knowledge of English, with TOEFL (Test of English as a Foreign Language) scores in the range of 500-575. Table 3.12 shows a summary of the subjects and the research design by native language, background knowledge of science, and language register of the text to be read.

The hypothesis was tested for each of three variables:

1. Language knowledge
   - LI--English as a first language
   - L2--English as a second language

2. Background knowledge of the subject being read
   - Expert--Scientific background
   - Nonexpert--Nonscientific background

3. Syntactic structure of the text being read
   - ST--Science and Technology
   - CL--Common Language
### Table 3.12

**Number and Distribution of Subjects Reading a Fiction Text**

<table>
<thead>
<tr>
<th>Language Background:</th>
<th>L1</th>
<th>L2</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Register:</td>
<td>ST</td>
<td>CL</td>
<td>ST</td>
</tr>
<tr>
<td>Background Knowledge:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experts (Scientists)</td>
<td>14</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Nonexperts (Nonscientists)</td>
<td>26</td>
<td>24</td>
<td>12</td>
</tr>
</tbody>
</table>

n = 76  n = 46  n = 122

ST = Syntactically Altered  
CL = Authentic

The reading comprehension scores for all of the subject groups were analyzed using multiple regression. The results of the analysis are shown in Table 3.13. In this study, neither the language register of the text—authentic (CL) versus syntactically rewritten (ST)—nor the background knowledge of the subjects—expert in science versus nonexpert in science—had a significant effect on reading comprehension scores. However, the effect of language knowledge—L1 versus L2—had a highly significant effect on both reading comprehension scores (p < .005) and on reading times (p < .0001). There was also a significant interaction effect when analyzing language knowledge and language register of the text together (p < .005); however, language knowledge was the only significant factor.
Table 3.13

Multiple Regression Results for Experts (Scientists) vs. Nonexperts (Nonscientists)

<table>
<thead>
<tr>
<th>Main Effects of:</th>
<th>on Comprehension Scores</th>
<th>on Reading Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR: Language Register of the Text: Authentic vs. Syntactically Adapted</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>BK: Background Knowledge:</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>LK: Language Knowledge: Native vs. Nonnative English</td>
<td>p &lt; .005**</td>
<td>p &lt; .0001**</td>
</tr>
</tbody>
</table>

Interactions

<table>
<thead>
<tr>
<th>Interaction</th>
<th>on Comprehension Scores</th>
<th>on Reading Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LK x BK</td>
<td>N.S.</td>
<td>N.S.</td>
</tr>
<tr>
<td>LK x LR</td>
<td>N.S.</td>
<td>p &lt; .005**</td>
</tr>
</tbody>
</table>

* Significant
** Highly Significant
Table 3.14 shows the mean comprehension scores for all subjects. It can be seen that the L1 subjects did score higher overall than did the L2 subjects. The one anomalous result was a slightly higher mean score by L2 than by L1 nonexperts reading ST syntax. In the other categories, scores of L1 readers were consistently higher than those of L2 readers.

Table 3.14 also shows the mean reading times (in minutes) for the subjects to read the text and complete the reading comprehension test. It is easy to see that the L2 readers took a significantly longer time (p < .0001) to complete the task than it took L1 readers, an average of more than twice as long. Several observations are interesting. Within the L1 group, the fastest readers were experts reading CL syntax. Within the L2 group, while the experts were also fastest, it was only within the ST syntax. When experts read the authentic fiction text (in CL syntax), the time for this group of readers was dramatically longer.

In summary, the statistical analysis showed that there were no significant differences among various groups except when controlling for language background. Neither the background knowledge of the subject to be tested or the register of the text itself had any significant effect on the subject's ability to comprehend what was being read. In this case, results indicate that readers are not troubled by all kinds of syntactic variants that have usually been assumed to be more complex and, therefore, interfere with comprehension: nominalizations, passives, and participle constructions. In the experiment, the revisions from a supposedly more simple common language register into a supposedly more complex science and technology register did not affect either the comprehension scores nor the reading time within each language group.
Table 3.14

Mean Reading Comprehension Scores and Reading Times for the Fiction Text

<table>
<thead>
<tr>
<th>Register of Text</th>
<th>L1</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expert (Scientist)</td>
<td>Nonexpert (Nonscientist)</td>
</tr>
<tr>
<td>ST</td>
<td>8.6</td>
<td>8.8</td>
</tr>
<tr>
<td>CL</td>
<td>8.2</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Mean Reading Score (out of 10)

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th>L2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Reading Time (in minutes)</td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>6.6</td>
<td>12.3</td>
</tr>
<tr>
<td>CL</td>
<td>5.8</td>
<td>19.3</td>
</tr>
</tbody>
</table>

\( n = 122 \)
3.5 Comments on the Procedures

One of the main limitations of each of the above studies is the amount of revising and way in which the sentences have been rewritten since that, of course, influences the results. As was the case with the writing studies discussed in Chapter II, the possibility of developing artificial texts through rewriting existed.

The framework of an authentic CS text or short story did not allow varying the syntactic features of the ST register in a very systematic way or making a valid check on item difficulty. To avoid artificial item descriptions, the sentences were used exactly as they appeared in the original text. Therefore, it would not have been reasonable to verify the reliability of our procedure by studying individual items or scores or by using new techniques such as Rasch scaling (Pumfrey, 1987). The adopted procedure was considered to be adequate for the research questions.

To make the study valid, a strong distinction had to be made between ST and CL syntax in each of the texts. However, if every sentence had been rewritten and thus the distinction had become even stronger, perhaps the results would have been biased because of artificial syntactic frequency loads.

The use of true/false questions for reading comprehension may have been another limitation. Since reading comprehension is not a phenomenon which is directly observable, it can only be measured indirectly, by asking some kind of question after the comprehension has taken place. True/false questions minimize the introduction of other variables, for example, skill at writing answers to questions.

Another limitation may be a ceiling effect of the reading comprehension scores. Since all of the subjects were able to obtain rather high scores, there is a strong possibility that the questions did not indicate as strong group differences as might have been possible with some other test instrument.
CHAPTER IV
CONCLUSIONS AND IMPLICATIONS

Some important questions have been raised by all of the writing and reading studies reported in this book:

Is there a difference in the way ST texts are written as compared with the way texts on a nonscientific subject are written? If a difference exists, is it a function of the writer—e.g., cognitive system such as background knowledge of the subject being written about (domain knowledge), language knowledge—and/or is it a function of the text topic itself—the subject matter being written about?

Is there a difference in the way ST texts are read compared with the way common language texts are read? If a difference exists, is it a function of the reader—e.g., cognitive system, background knowledge of the subject being read, language knowledge—and/or is it a function of the text—e.g., syntax, lexis, discourse structure?

Since writers must write text which is to be comprehended by specific readers, and conversely, since readers must be able to comprehend what writers produce, the answers to the above questions, as well as to any interactions between them, are critical to the field of EST.

4.1 Conclusions of the Writing Studies

Table 4.1 gives an overview of the significant results of all of the writing studies discussed in this dissertation.

In the original study (FIT/TUE 1), subjects were asked to complete sentences within a computer science context. Although the question of the influence of the surrounding ST cotext was raised, the results indicated that both L1 and L2 writers, those with or without computer science background, preferred the use of ST syntax within the ST text. However,
Table 4.1

Summary of Results of the Effect of Specific Variables on a Writer's Choice of Syntax

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT/TUE 1</td>
<td>N.S.</td>
<td>Significant</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>FIT/TUE 2</td>
<td>N.S.</td>
<td>N.S.</td>
<td>Highly Significant</td>
<td>N.A.</td>
</tr>
<tr>
<td>FIT/TUE 3</td>
<td>Highly Significant</td>
<td>N.S.</td>
<td>Highly Significant</td>
<td>N.S.</td>
</tr>
<tr>
<td>Paragraph</td>
<td>Significant</td>
<td>N.S.</td>
<td>Highly Significant</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

N.S. = Not Significant  
N.A. = Not Applicable

Note: See individual studies for specific results
background knowledge was a factor since CS subjects used significantly more ST syntax within the ST text than humanities subjects did.

In FIT/TUE 2, using both a computer science text and a fiction text, the results were different. Both expert and nonexpert, native and nonnative writers preferred to use CL syntax within the fiction context and ST syntax within the computer science context, even though the surrounding cotext for both articles was written in CL syntax. In this study, the possible influence of the surrounding cotext was eliminated for the computer science article, but was raised for the fiction text. Since that text was already written in the CL register, and since all groups preferred writing in the CL register within that text, was there a possible priming effect?

The third study, FIT/TUE 3, was considerably refined. It was designed to replicate the premise of the previous studies but was enhanced to eliminate a possible learning effect from the cotext in both the computer science (ST) and fiction (CL) texts. Both texts were produced in an ST syntactic register and in a CL syntactic register. Half of each group (L1 vs. L2, expert vs. nonexpert) worked with both texts in the ST register; the other half worked with both texts in the CL register.

The results of this third study are quite interesting, especially compared with the previous two studies. The hypothesis that the text topic itself, rather than either the influence of the surrounding cotext or knowledge of the topic of the text, is the main factor which determines the syntax chosen by the writer was strongly supported.

While the Dutch, as well as the American writers, clearly differentiated between syntax use in the fiction and computer science texts, the Dutch used more ST syntax in both texts than did the Americans. In looking at within group differences, it is interesting to note that the Dutch humanities subjects used slightly more ST syntax in an ST cotext than did
Dutch computer science subjects; there was a slight reverse effect in the CL cotext where computer science subjects used more ST syntax than humanities subjects.

For both the Dutch and American subjects, the background knowledge of the topic being written about (e.g., computer science majors writing about computer science) had no significant influence on syntactic choices. In addition, the cotext (ST vs. CL) did not have a significant effect; however, for the Dutch, in contrast to the previous studies, the linguistic cotext did have a slight, though not significant, effect. They evidently experienced some priming effect from the surrounding text when they were deciding which syntactic structures to use.

The original question still stands: what affects a writer's choice of syntax within a particular text? The fourth study attempted to get closer to a more definitive answer by completely removing the issue of the possible influence of surrounding cotext.

Thus, in the final writing study, all subjects were given a free writing assignment with no cotext to provide any kind of influence. Each subject had to write a paragraph on the scientific subject of gravity and another paragraph on the nonscientific subject of newspapers, in each case using only a word list as a cue. Again, all subjects significantly preferred the use of ST syntax in the scientific context and CL syntax in the nonscientific context, providing strong support for the hypothesis that syntactic choice is text dependent—a function of the text topic or context in which the writer is working. Therefore, this primary hypothesis was strongly supported in all of the studies.

How was the choice of syntax affected by the other variables in these studies, namely, background language (L1 vs. L2) and background knowledge of the topic being written about (expert vs. nonexpert)? The hypothesis that there would be a significant difference in the number of ST
syntactic structures chosen by L1 subjects as opposed to L2 subjects was not supported in all of the FIT/TUE studies and there are, as a result, some interesting observations which can be made. In the first two FIT/TUE studies, there was not a significant difference in ST syntax use between the American and Dutch subjects; however, in the last FIT/TUE study and the paragraph writing study, there were significant differences between L1 and L2 language groups. In FIT/TUE 3, the Dutch used significantly more ST syntax in both the computer science and the fiction texts than did the American subjects. However, in the final writing study, in which subjects wrote paragraphs with no surrounding cotext, L1 writers used significantly more ST syntactic structures than did L2 writers (in this case, mixed ESL) in both paragraphs, but especially within the paragraph about the scientific topic.

The explanation for these differing results on the effect of language background on choice of syntax may depend on differing second language backgrounds of the subject groups. In the FIT/TUE studies, the first language of the subjects was Dutch and they were operating in a nonEnglish-speaking environment. Their preference for ST structures may come from frequent exposure to English texts, many of which are scientific and/or technical texts. Since much of what they read, especially for academic purposes and especially at the advanced levels of these Dutch subjects, is in the ST register, it is not surprising that their writing would, in many cases, imitate that ST form. For the last study, however, the subjects were mixed ESL, 31% of whom were Arabic. While they were in an English-speaking environment at a US university (FIT), since their average length of stay in the US was slightly less than five months, it is not believed that the English-speaking environment was a strong influence for this group of subjects. In searching for other influencing factors, it is highly possible that, although their level of English was also advanced, their
exposure to common language materials in their ESL/EFL classes was greater than to scientific materials. It would, of course, take a great deal of study to verify these suppositions within each language group. Contrastive linguistic analysis, which might uncover other influencing factors, is beyond the scope of this book.

The third hypothesis, that there would be a significant difference in the number of ST syntactic structures chosen by experts in computer science or by subjects with a strong scientific background as opposed to the number chosen by nonexperts or subjects without a strong scientific background was not supported. This finding is in line with all of the studies except the first study in which expert subjects preferred to use ST structures within an ST context, all of which was written in the ST register. Thus, the conclusion can be supported that having a scientific background does not cause a writer to use significantly more scientific (ST) syntactic structures than a writer without such a background.

As was discussed in Chapter I, there seem to be some definite reasons for choosing certain syntactic structures within the expository text of the ST register, usually dictated by the nature of the subject matter itself. The four syntactic structures chosen for this study—infinitives, nominalizations, participles, and passives—are often used to make the actor secondary to the action and to present scientific and technical information in an impersonal way, an essential feature of the ST register.

4.2 Conclusions of the Reading Studies

An overview of the significant results of all of the reading studies discussed here is presented in Table 4.2.

As was discussed in Chapter III, the three FIT/TUE writing studies included a reading component in which subjects were given words or groups of words, presented in alphabetical order (violated syntax), within a
Table 4.2

Summary of Results of the Effect of Syntax on Reading Comprehension Scores

<table>
<thead>
<tr>
<th>Study</th>
<th>Language L1 vs. L2</th>
<th>Background Knowledge Expert vs. Nonexpert</th>
<th>Text Topic Scientific vs. Nonscientific</th>
<th>Linguistic Context ST vs. CL</th>
<th>Time L1 vs. L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIT/TUE 1</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.A.</td>
<td>Significant</td>
</tr>
<tr>
<td>FIT/TUE 2</td>
<td>Highly Significant</td>
<td>Significant</td>
<td>N.S.</td>
<td>N.A.</td>
<td>Significant</td>
</tr>
<tr>
<td>FIT/TUE 3</td>
<td>Highly Significant</td>
<td>Significant</td>
<td>N.S.</td>
<td>N.A.</td>
<td>Highly Significant</td>
</tr>
<tr>
<td>Reading CS</td>
<td>Highly Significant</td>
<td>Significant</td>
<td>N.S.</td>
<td>N.S.</td>
<td>Highly Significant</td>
</tr>
<tr>
<td>Reading Fiction</td>
<td>Highly Significant</td>
<td>N.S.</td>
<td>N.S.</td>
<td>N.S.</td>
<td>Highly Significant</td>
</tr>
</tbody>
</table>

N.S. = Not Significant
N.A. = Not Applicable

*See individual studies for specific results
text. Thus, subjects in these three studies were first solving the problem of putting the sentence elements in order to create meaning and then reading the results. In all three of these studies, subjects answered 10 true/false questions based on the 10 sentences in the text which were written with violated syntax. The same basic questions were addressed in each of the three tests; namely (1) what effect does the background knowledge of the subject (expert vs. nonexpert) have on reading comprehension? (2) what effect does the language knowledge of the subject (L1 vs. L2) have on reading comprehension? and for the last two studies, (3) what effect does the topic of the text (computer science vs. fiction) have on reading comprehension? and (4) what effect does the syntax of the surrounding cotext (ST vs. CL) have on reading comprehension?

The first study, FIT/TUE 1, had all subjects read an excerpt from a computer science journal article in which the 10 sentences were surrounded by cotext written in the ST (syntactic) register. A statistical analysis showed that there were no significant differences in comprehension among the groups, whether controlling for language background, background knowledge of the subject being read, or register of the text itself. An interesting finding is that presenting some sentences with violated syntax within a text did not significantly hinder comprehension. While all groups showed some differences between nonviolated and violated syntax in both ST and CL registers, the Dutch humanities students showed the greatest loss of reading comprehension as they attempted to solve the problem of dealing with violated syntax.

In FIT/TUE 2, subjects read violated syntax within both the computer science text, which had the surrounding cotext rewritten in the CL register, and the fiction text, which was already in the CL register. The results confirmed that again the syntax did not have a significant effect on reading comprehension for any of the subject groups. However, there
were significant differences in reading comprehension between the L1 and L2 subjects and between the expert and nonexpert readers, with L1 experts in computer science having the highest overall reading comprehension scores.

FIT/TUE 3 had half of each group of subjects (L1 vs. L2, expert vs. nonexpert) read both the CS and the fiction texts in the ST register and half read both texts in the CL register in an attempt to eliminate the question of the influence of the surrounding context on overall comprehension. Again, the sentences they were being tested on were written in violated syntax; therefore, the subjects had a problem-solving task of constructing meaning from syntactic elements presented in alphabetical order. Thus, in this study, as in FIT/TUE 1 and 2, the syntax of the sentences presented in a violated form was not being manipulated; rather, the syntax of the surrounding context was being manipulated. It is not surprising that in this study too, the syntax of the surrounding context did not have a significant effect on reading comprehension scores. It is particularly interesting that the violated syntax proved little hindrance to reading comprehension for all of the subject groups in all three tests.

The results indicated that there was a significant effect of background knowledge on comprehension, that having a background in computer science improved comprehension of the CS text. While there is a question of a possible ceiling effect since all scores were quite high (8.4 out of 10 for the Americans and 6.7 out of 10 for the Dutch), at least for this group of subjects, the American CS majors were the best overall readers. In addition to scoring higher on the CS text, they scored higher on the fiction text than did the American humanities subjects. Recall that this was also true in FIT/TUE 2. There was also a significant difference between L1 and L2 readers, not surprising since the American subjects had the advantage of reading in their native language while the Dutch had the
disadvantage of reading in a second language.

Finally for this study, there was no significant overall effect of text topic (fiction vs. computer science) so both texts were equally difficult to understand for this group of readers. However, for both American and Dutch humanities subjects, the difference in reading comprehension scores on the CS text was not great because both groups, regardless of language background, lacked knowledge of computer science.

In addition to the three studies reported above which were primarily designed as writing tests but which included a reading section, there were two additional reading studies reported herein. The first was also a cooperative effort between Florida Institute of Technology (FIT) and Eindhoven University of Technology (TUE); the second was done exclusively at Florida Institute of Technology. Both studies had the same basic research design and addressed the same basic research questions. To what extent do the following factors affect reading comprehension: (1) syntactic rewriting of a text from ST to CL syntax (considered by some to be a simplification); (2) language background of the reader (L1 vs. L2); and (3) background knowledge of the reader (expert vs. nonexpert)?
syntax and for this test, an alternate version was developed by writing it into ST syntax. As in the first study, the syntax of the text (ST vs. CL) did not significantly affect reading comprehension scores for the subject groups. In addition, for this fiction text, there were no significant differences in scores between readers with a scientific vs. nonscientific background. However, as in the previous studies, L1 readers did score significantly higher than L2 readers, again showing the advantage of reading in the native language as opposed to reading in a second language.

4.3 The Effect of Time In the Reading and Writing Tests

Reading times were recorded for each of the writing and reading studies and indicate time to complete the entire test. While they are not perfectly reliable since they were recorded by the subjects themselves, they indicate some interesting conclusions. (See Table 4.2 for a summary.) It is certainly not surprising that in each of the studies, L1 subjects reading their native language completed the test in significantly shorter time than did the L2 subjects who were reading a second language.

When looking at within group differences in the individual studies, some of the findings are interesting, yet they are not consistent among the studies. Recall that in the FIT/TUE studies, subjects had the problem-solving task of reading violated syntax within surrounding context. In FIT/TUE 1, the CS experts tended to take more time than nonexperts (humanities majors) to read the CS text. However, in both FIT/TUE 2 and FIT/TUE 3, where subjects read both the CS and fiction texts, the nonexperts took longer to complete the task, with the difference being a significant one for the Americans.

The two separate reading studies (which had subjects read text with normal, nonviolated syntax) yielded slightly different results. As with
the FIT/TUE studies, the L2 subjects took significantly longer to complete the test than did the L1 subjects. In the first study, where subjects read either an authentic or syntactically altered (ST to CL) version of the computer science text, time differences were examined for the two versions. Within the groups, for all subjects except the L1 nonexperts, the ST version took slightly longer to read than the CL version. In the second study, where subjects read either an authentic or syntactically altered (from CL to ST) version of the fiction text, again the ST version took slightly longer to complete, except for all of the L2 subjects.

There seem to be a few general conclusions which can be reached from the information on time to complete the test. Certainly, a subject reading in a second language takes longer to process and comprehend the text. In some cases, it appears that experts reading in their field may take longer because they are using their background knowledge to get at a deeper meaning of the text than those without such knowledge. And finally, subjects take somewhat longer to process text written in ST syntax than that in CL syntax, although their final comprehension scores on the two texts are not significantly different. Therefore, an important conclusion seems to be that the complexity of the syntax affects time, but not comprehension.

4.4 General Conclusions

In the beginning of this book, some important questions were raised, questions to which there are now at least limited answers. For the writing studies, the primary questions dealt with what causes a certain writer to choose particular syntactic structures within a certain context or topic. The overall conclusion to be reached is that all subjects--L1 and L2, expert and nonexpert--prefer to use ST syntax in scientific texts and CL syntax in nonscientific texts. The issue of the influence of any surrounding
cotext on syntactic choices was eliminated for most subjects. These studies give strong evidence that writers have an equal option of choosing either the ST or the CL register in syntactic structuring yet the text topic itself is the primary factor in determining a writer's choice of syntax, regardless of that writer's language background or expertise in the subject matter being written. Thus, at least for the syntactic element, the ST register is a psycholinguistic reality.

For the ST register, the topic of the text has a stronger impact on ST syntax use than background knowledge or language knowledge. The text's subject automatically suggests the syntax to be used regardless of the writer's background knowledge and language knowledge. Writers tend to tune their linguistic register to the topic in a psycholinguistic way; therefore, the ST register is not merely a matter of linguistic description related to linguistic cotext, background knowledge, or language knowledge. Of these factors, language knowledge is the only one shown to have an occasional impact. Background knowledge rarely impacted the results, and there was no effect of the linguistic cotext on the writers' syntactic choices.

Of course, the results cannot be overgeneralized. This series of research studies dealt solely with syntactic choices, and, in this case, only with four specific syntactic structures common to the ST register--infinitives, nominalizations, participles, and passives. Thus, the results cannot be generalized to the entire syntax of the ST register.

In addition, as was discussed earlier, the ST register is concerned with more than just syntax. Scientific and technical lexis or terminology is another matter, and one that could vary greatly within each LSP (Languages for Specific Purposes) field. (See for example, Ulijn, 1981, for a discussion of foreign language reading, and Lankamp, 1989, for a discussion of the issue of medical terminology.) Whereas background
knowledge has hardly any effect on the syntax chosen by the writer, it would have a strong effect on word choice as has been convincingly demonstrated in experiments by Lankamp. As discussed in Chapter 1, such properties of the register as discourse structures must also be considered. Further research on the influence of background knowledge and language knowledge on discourse structures used by certain writers within certain ST texts would perhaps give more enlightenment regarding the psycholinguistic base for that register within the writer.

For reading, the primary question was whether the ST register, with its supposedly more difficult structures, such as passives, nominalizations, participles, and infinitives, is more difficult to comprehend than the CL register. These reading studies strongly concluded that the syntax of the ST register did not cause more comprehension difficulties than that of the CL register for any of the subject groups. In addition, both L1 and L2 readers do not seem to have their comprehension greatly hampered by the amount of violated syntax (sentence elements presented in alphabetical order) used in these studies. This data supports the idea that in a conceptually guided and partially parallel strategy for processing texts (discussed in Chapter III), a thorough syntactic analysis is not needed either by native or nonnative readers.

As discussed above for writers, the ST register is certainly more than syntax, and for readers, the same factors such as lexis and discourse structures are elements whose effects on reading comprehension must be considered.

4.5 The Interactive Reader-Writer Readability Model

Now we can return to the Interactive Reader-Writer Readability Model introduced in Chapter I and shown in Figure 4.1. The results of previous readability studies yielded varying conclusions. Early readability
Figure 4.1 An Interactive Reader-Writer Readability Model
formulas dealt with mainly surface level textual features and thus the validity of their results has been questioned. In addition, it has been concluded that the applicability of such formulas to guide the way writers write is very limited. (See Chapter I for a complete discussion of this research.)

What is needed is a deeper insight into the cognitive processes within the reader and the writer and insight into how these are interrelated. Since the research presented in this book has been limited to the syntax of the ST register and how it affects both the reader and the writer, the Interactive Reader-Writer Model is meant to demonstrate only that part of the system.

In many readability studies (and in the formulas that evolve from some of those studies), the analysis of the textual features is the primary task; following from that is an analysis of how those particular features affect the reader. Therefore, if any relationship has been defined as part of the concept of readability, it is a relationship between the reader and the text. A very important element which is missing from many readability studies is the direct relationship of the writer to the text and the resulting indirect relationship of the writer to the reader.

This Interactive Reader-Writer Readability Model, using the technique of Venn diagrams, seeks to incorporate all of these critical elements--writer, reader, and text--in an interdependent relationship incorporating the factors which influence each element. The writer, as the generator of the text, is examined first. The current research examined the effects of background knowledge, language knowledge, and linguistic cotext on a writer's syntactic choices. It was found that language knowledge was significant in two of the four studies, background knowledge in only one of the four, and linguistic cotext in none of the studies. However, a very important finding was that the topic of the text
itself always had a significant effect on the kind of syntax used by the writers. The ST register, including its syntactic elements, therefore, resides within the writer's psycholinguistic system and is called forth when triggered by a particular text topic. The effects of background knowledge and language knowledge, to the extent they exist, do not contradict this psycholinguistic system. The writer chooses her language register as a function of the topic with language knowledge and background knowledge occasionally becoming influencing factors.

These syntactic elements which have been used by the writer become an inherent part of the text itself. The reader then interacts with the text, bringing to it his own psycholinguistic system. When the current research examined specific elements which affected readability of the text, neither the text topic nor the linguistic register was a significant factor in reading comprehension for all of the subject groups. However, the domain-specific background knowledge of the text topic was a significant factor in three of the four studies which included a scientific (CS) text. Moreover, language knowledge is almost always a significant factor in both reading comprehension and time to process the text. Thus, the elements within the reader strongly influence how he or she engages the interaction between these psycholinguistic aspects and the readability of the text.

4.6 Practical Implications and Suggestions for Further Research

Since writers tend to tune their linguistic register to the topic of the text, independent of background knowledge and language knowledge, technical writing is, for both natives and nonnatives, not only a matter of language and specialist knowledge training, but also of gaining insight into the psycholinguistic aspects of writing. Therefore, researchers must
continue to analyze the psycholinguistics of writing and address such questions as: If the ST and CL registers are systematized in the writer's brain, how are these systems initially developed? What are the parameters of these systems? How much and what kind of knowledge must a writer have to encode an ST register? What level of second language proficiency must be reached before the concept of register takes effect? Does this depend on a writer's first language to determine the amount of linguistic transfer?

The role of background knowledge in both reading and writing needs much deeper analysis. The current studies have clearly shown that text topic determines syntactic choice. However, a nonexpert may choose ST syntax within a scientific topic, write a perfectly grammatical sentence, yet produce a sentence that is inaccurate or illogical because of lack of background knowledge of the topic. That situation occurred in the studies reported in this book.

As part of the FIT/TUE 1 study, an error analysis was performed to itemize and then compare the types of errors made by both L1 and L2 students. (See Strother & Ulijn, 1989a, for complete details.) Because of the nature of the task, the number and kinds of errors were limited in each sentence, but the results were still striking. In general, it could be concluded that L2 subjects made more errors than did L1 subjects. However, it seems that background knowledge of the topic being written about seems to have a stronger impact on number and kind of errors than language background of the writer. For example, in one of the sentences within the CS text, subjects should have written, "... the problem is separating and clustering sets of files and providing a mechanism. . . ." A number of subjects wrote the following grammatical ST sentence which contained a logic error: "... the problem is separating clustered sets of files. . . ." This reflects, not a problem with syntax, but rather a problem
stemming from lack of knowledge about the topic. This qualifies our results that while background knowledge rarely affects the syntactic choices of a writer, it is still a critically important factor in writing in science and technology, in particular in peer writing in specialists fields.

For second language writers, there can be a double hurdle—difficulty with the second language and lack of knowledge about the topic. Ideally, therefore, one should use a native writer with strong background knowledge for technical writing tasks. In situations where this "ideal technical writer" is not available, a serious question is faced by business and industry; namely, should one hire technical writers who are native speakers but who might not have background knowledge of the topics to be written, or technical writers who are nonnative speakers but who do have the relevant background knowledge? Can industry better train a skilled, nonspecialist writer (e.g., a linguist) in the necessary domain-specific knowledge or better train a specialist in the field (e.g., an electrical engineer) to become a skilled writer? At what level does second language knowledge become adequate to deal with the above issues?

These questions have no easy answers; however, the current research at least gives some general direction and then leaves the field open for much-needed further research. Having native speakers or second language speakers with near-native proficiency eliminates two of the most serious language problems faced by a second language writer with low L2 proficiency in a technical writing situation: (1) because of the reading-writing interaction, the lower the L2 proficiency, the more difficulty a writer has in reading the specialist materials with which to work; (2) writing in a second language often causes problems within the text, thereby lowering the readability of that text.

From the evidence culled from the systematic error analysis of FIT/TUE 1 and the superficial error analysis of the other studies, it seems
clear that lack of background knowledge of the topic causes errors to occur that would completely misrepresent scientific or technical information. As stated above, these errors usually had little or nothing to do with syntax and were often present within grammatical ST structures. Therefore, background knowledge seems to be a strong criteria for technical writers, and again, there is the writing-reading connection. Perhaps reading to acquire background knowledge of the field to be written about may be the most critical task for improving technical writing. And improved writing increases readability.

The question of readability is one that also must continue to be addressed. As was discussed in Chapter 1, many of the current readability formulas are measurements of certain characteristics of a text but are not useful in predicting how difficult a specific text is going to be for a specific reader. Thus, readability formulas should not be used as guidelines for technical writers.

The question of readability also has to address the issue of audience. For example, should a technical writer change the syntax used depending on whether she is writing for a specialist or for a lay reader? The current research makes it clear that this is not necessary. Since reading comprehension scores were not significantly affected by ST versus CL syntax, audience should not determine the syntactic structures used, assuming the readers share similar characteristics with the subjects in these studies. However, it may be true that syntactic simplification may have an effect on groups other than those tested in this experiment, such as limited proficiency language users (see for instance, Beaumont, 1982, who found syntactic problems in early L1, English, reading by children).

Of course, other factors not taken into account by these studies, such as lexis and discourse structure, could have a significant influence on the readability of a technical text by lay versus specialist readers. Huckin

106
(1983) points out that, using protocol analysis, cognitive scientists have discovered that experts process information differently than novices. If writing for experts, cues should be included that trigger, in long-term memory, the experience and knowledge needed to understand the text. If writing for novices, analogies from common knowledge can be used to help the readers acquire new knowledge which eventually becomes part of their own expert systems as new knowledge accumulates and is called forth in a variety of applications. The basic message of cognitive research is that if the thought-processing patterns to be used by the reader are identified, the writer can design the text to facilitate that kind of processing (Samuels, 1988). Schema theory applies then not only to reading but also to writing.

Focus also has to kept on current trends in technical writing. While the current research does not support the need to do so, many technical communication and business communication textbooks (e.g., Rew, 1989; Andrews & Andrews, 1988; Blicq, 1986; Bovee, 1986; Markle, 1988) strongly encourage use of CL rather than ST syntactic structures in many business and technical applications. If this tendency becomes more widely adopted by the industrial community, it will be interesting to see whether or not writers' syntactic choices within different contexts will shift, in particular what will happen to peer writing in the highly specialized fields of science and technology where the action is more important than the actor and ST syntax is often more efficient and less ambiguous.

4.7 Teaching Implications

In considering the interrelationship between reading and writing, we must determine how the research findings of this book can be incorporated into LSP curricula and technical writing and reading situations in both universities and business and industry. Both language
knowledge and background knowledge have been more or less eliminated as affecting factors in syntactic choice within registers, but, of course, in both instances, a certain level of knowledge is assumed—in this case, advanced levels of both English and computer science or some other scientific field. It is yet to be determined what level is necessary to assume that both the language knowledge—either first or second—and the background knowledge of the field are sufficient for the register to be in place in the brain.

Therefore, overriding questions still seem to be: do we need to teach an ST register in any university or industrial setting? Should specific syntactic structures be taught within certain contexts depending on learner need? How can the reading-writing connection be improved in pedagogic situations?

Hutchinson and Waters (1987) claim that there is no such thing as ESP methodology, that ESP is not different from any global approach to improved writing skills. However, while the methodology may not be extremely different, the content certainly will be. For example, students should be helped with features identified as typical within a particular context of use and which the student is likely to face in a target environment. This calls for specific instruction of these particular features. Smits and Ulijn (1990) state that if one wants to achieve real proficiency in special purpose language at a professional level, grammar has to have a prominent place in the curriculum. The results of the current studies call for a modification of that view, limiting the study of syntax to those specific features that cause specific writers difficulty or stressing specific syntactic features for specific purposes.

For example, teachers could help students see advantages of the active verb form—being more forceful and/or more personal, assigning responsibility—and the advantages of the passive verb form—emphasizing
receiver or results of an action, avoiding first person, avoiding assignment of responsibility, and avoiding a long phrase as the subject. The key, of course, is helping students see the appropriateness of both verb forms in given situations and helping them determine the focus to be applied in a specific context for a specific audience. In science and technology, it is often the passive which carries the appropriate focus and students need to know that.

Instead of focusing on syntax, the focus should be on development of the previously discussed textual skills in obtaining the required information from the texts and on the development of both the technical and subtechnical vocabulary. In this respect, discourse analysis would be very useful in diagnosing difficulty in L2 reading comprehension, as has been shown by Bensoussan (1987). Readers should be guided toward focusing more on concepts and on the vocabulary, including that related to their background knowledge, with syntactic analysis being superficial in most cases. Of course, since lexical and textual simplification were not tested in the present studies, these aspects call for detailed verification before drawing any final conclusions about their simplifying effects.

Second language professors have to help students bring their global skill levels up to a point where they can write successfully in the target language. Then these writers can focus on the specific elements of the ST register so they can move from general language writing to skillful technical writing.

In addition to the above, if an awareness of the reader's needs must be communicated to the writer and an awareness of the writer's purpose to the reader, then that too must be taught. The strategy of teaching reading and writing together can help achieve those goals. As Selinker (1984) observed, in specific purpose (here ST) language acquisition, language and subject matter are intimately and structurally
Chapman (1984, 1987), for example, says that if early attention is called to the way meaning is constructed in texts, students become aware of cohesive ties and the different indices of register even at very early stages of reading and writing development. As MacLean and Chapman (1989) point out, children develop rudiments of register during their early socialization in school by becoming sensitive to language variation. By encouraging this appreciation of audience in all forms of language, the foundations can be laid for the later intricacies of subject area register. They believe that by making students more aware of cohesion and register throughout primary and secondary grades, they will become better writers and readers.

Trimble (1985, p 14) confirms this by stating that "we have found writing best approached as a transfer technique. That is, we have the students consciously practice the rhetorical concepts they have found in their reading by giving them writing exercises designed to make them choose those rhetorical elements most appropriate for a given purpose and given level of reader. In requiring students to choose specific rhetorical functions and techniques for the presentation of their EST information, we also strengthen their recognition of these functions and techniques when they read EST discourse."

Certainly, if this is done, the process of text analysis for both readers and writers should continue through university education and industrial training programs, especially for specialists in a particular field who must write readable text for both lay and specialist readers. If the quantity and quality of text comprehension is enhanced, then texts should become more accessible to all kinds of readers. Thus, concern with the logical organization of a text is as much a reading as a writing problem. The more we can help students focus on this fact, the higher the quality of output, whether that output be written text or text comprehension.
Since writing is a process through which meaning is created, the writer must have background knowledge of the content to create meaning in a particular field. As stated earlier, reading to acquire background knowledge of the field to be written about may be the most critical task for improving technical writing. So again, perhaps especially in specialist fields, reading and writing are strongly interrelated. Therefore, as Krashen (1982, 1987) has pointed out, improved reading can improve writing output. Krashen also predicts that the use of specialized reading materials, because they generate high reader interest, improve reading comprehension, and in general, further language acquisition.

The EST teacher, whether in a university classroom or an industrial teaching situation, should make available in the classroom large quantities of interesting reading material in the learners' content areas. Since teachers often face groups of students with heterogeneous backgrounds, whenever possible, individualized instruction helps students make full use of their background knowledge and/or allows them to develop such knowledge.

Since these studies have demonstrated that syntactic simplification does not enhance comprehension, the use of authentic text in second language reading instruction has also been confirmed. The author (Strother, 1987) has directly applied the above findings and implications in the writing of an EST/EAP textbook for advanced second language learners in which authentic selections from university textbooks and reading materials are used. The reading selections are in the university students' major fields, so reader interest is high, motivation is high, and reading comprehension improves.

This brings us full circle back to the question of whether to teach EST, and if so how. The answer is perhaps a qualified yes. The teaching of EST should be an enhancement of general language skills aimed at
reaching specific learners. There must be a definite focus on the interrelationship of the reader, the writer, and the text. Since there is a very definite ST register, the more the reader knows about it, the better he should be able to comprehend what is written in that register. The more the writer knows about the register and how a reader deals with it, the better she will be able to produce high quality text.
APPENDICES
APPENDIX A

Test Booklets

READING COMPREHENSION AND WRITING TEST
Science/Technology Version

Computer Science Text—Authentic Version

Fiction Text—Syntactically Altered Version
READING COMPREHENSION AND WRITING TEST

NAME______________________________

MAJOR______________________________

NATIVE LANGUAGE____________________

AGE__________ CIRCLE ONE: FRESHMAN
SOPHOMORE
JUNIOR
SENIOR
GRADUATE STUDENT

STARTING TIME:__________

ENDING TIME:__________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
INSTRUCTIONS

Read the given texts carefully in order to answer the True/False questions afterwards. When you start reading the texts, record your beginning time on the cover sheet and when you have finished the texts, record the time again. Answer the True/False questions using the texts for reference.

Ten passages appear in italics within the texts. Within such a passage, words or groups of words are enclosed in parentheses ( ). The word or word groups are listed in alphabetic order. You are to make complete sentences of the passages in italics using the words in parentheses.

You may: - change the order of the groups of words; - change the forms of words (for example, adjective into adverb or verb into noun; - conjugate the verb; - add some words if you feel it is necessary to complete a sentence.

All the words or word groups within the ( ) have to be used. The word order within the ( ) must remain the same. For each ( . . . ) you may add any number of words you would like or you may choose not to add any words.

If there are alternative solutions for any sentence, write the one you thought of first.

Write the sentences on the answer sheet.

EXAMPLE

( . . . ) (computer simulations) (look) (use) the past and future of galaxies. The computers are programmed with data on the mass and size and relative locations of the galaxies.

POSSIBLE ANSWERS

1. Computer simulations are used to look into the past and future of galaxies.

2. Computer simulations can be used to look into the past and future of galaxies.

3. One can use computer simulations to look into the past and future of galaxies.
PART ONE

This text is concerned with computer science. It deals with the storage of great numbers of data in a computer.

MASS STORAGE SYSTEMS AND EVOLUTION OF DATA CENTER ARCHITECTURES

The investment that an enterprise makes in data processing equipment was once concentrated in the data center. The architecture of that data center is now changing into what might better be called a collection of cooperating subsystems that encompasses the geographic extent of the enterprise. (. . .) (connect) (remote terminals) (telephone lines, leased lines, concentrators, etc.) to a front-end processor which often permits access to one of several host processors at the data center.

Inside the data center, the multiple hosts may be supplied by a single vendor, although processors representative of several vendors are becoming more common. If the data center itself does not contain multivendor hosts, the enterprise may well face a problem similar to having multiple data centers, located in different areas, that have grown separately around products from a particular vendor. Sometimes the several vendors are represented for reasons that are lost in history, but the fact remains that users throughout an enterprise desire access to files and other services that are distributed over several host processors. Associated with these data centers is the need for a central data repository of some description, whether it is based on paper, microfilm, or magnetic tape.

The typical central data repository in the DP world is the magnetic tape library. By "central repository" we mean the place from which data volumes are brought for entry into the system. The general paradigm is that the tape reel is fetched from the repository and mounted on a drive. Then data are moved from the tape volumes to secondary storage (i.e., disks) for processing. Of course, once the reels are mounted, some data are processed directly from the magnetic tape because of the very short access time to the "next record." During the last decades, (. . .) (consider) the density of storage of the tape and the speed with which tape is moved, tape technology has become only a few hundred times
better. During that same period of time, considering the instructions executed per second and the increased power of an individual instruction, the host computer (i.e., the central processor and main memory) has increased in power several thousand times. This disparity in rate of improvement has placed great pressures on the system to "handle more tape faster."

(handle) more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day. This is a labor-intensive, error-prone operation. The problems of erroneous tape mounts and the associated time delays have led many managers to keep their data bases "on line" on spinning magnetic disks (i.e., secondary storage). The cost of doing this includes the 100:1 (or more) differential in media costs plus the costs of relatively large amounts of electrical power and floor space that are required by disk drives.

The data center manager who is concerned about (reduce) operating costs or (improve) performance of this central repository may introduce optical data disks. These disks have been anticipated for some time now by such authors as Kenny and Brodie. R & D in the area of optical data disks has been stimulated by the efforts in the video disk field of consumer electronics. At this time, however, it is not possible to purchase "off-the-shelf" devices for introduction into a data center.

Higher storage density and lower per-bit costs are promised by optical data disks which, however, tend to employ a write-once, read-many-times technology that does not lend itself to plug-compatible replacement at the device level. (at a higher system level) (current devices with optical data disks) (replace). The programmer must develop considerably more software to take advantage of optical data disks. Even though much magnetic tape is used in a write-once, read-many-times manner, there remains a psychological barrier to the introduction of a new technology that forces this mode of operation.
This text is from a short story about a man who is fighting against cold and the elements to survive.

From: TO BUILD A FIRE

Day had broken cold and gray, extremely cold and gray, when the man decided to turn aside from the main Yukon River trail and to climb the high earth-bank, where a vague and little-travelled trail led eastward through the timberland. The steepness of the bank caused him to pause for breath at the top, excusing the act to himself by looking at his watch. It was nine o'clock.

The man's look was flung back along the way he had come. The Yukon River lay a mile wide and was hidden under three feet of ice. This ice was covered by as many feet of snow. It was all pure white, rolling in gentle waves where the ice-jams of the freeze-up had formed. North and south, as far as could be seen by his eyes, it was unbroken white.

Turning to go on, he spat pensively. He was startled by a sharp, explosive crackle. He spat again. And again, in the air, before it could fall to the snow, the spittle cracked. He knew that spittle was made to crack in the air at fifty below zero. Undoubtedly it was colder than fifty below--how much colder was not known. But the temperature did not matter. He was bound for the old camp on the left fork of Henderson Creek, where the boys were already. They had come over from the Indian Creek country, while he had come the roundabout way to take a look at the possibilities of getting out logs from the islands in the Yukon in the spring. He would be in the camp by six o'clock; a little after dark, it was true, but the boys would be there, a fire would have been built, and a hot supper would have been prepared. As for lunch, he pressed his hand against the protruding bundle under his jacket. It was also under his shirt, wrapped up in a handkerchief and lying against the naked skin. It was the only way to keep the biscuits from freezing. He was smiling agreeably to himself as he thought of these biscuits; each had been cut open and sopped in bacon grease and a generous slice of fried bacon was enclosed in each.

In fact, nothing was being carried by him but the lunch wrapped in the handkerchief. (. . .) (cold) (surprise), however. His conclusion was that it certainly was cold, causing him to rub his numb nose and cheekbones with his mittened hand. He was a warm-whiskered man, but (. . .) (the hair on his face) (his high cheekbones and nose) (not) (protect).
The man was followed by a dog trotting at his heels, a big native wolf-dog, gray-coated and without any visible or temperamental difference from its brother, the wild wolf. The animal was depressed by the tremendous cold. Its knowledge told it that it was no time for travelling. It was seventy-five below zero. The freezing-point is thirty-two above zero. Nothing about thermometers was known by the dog. But the brute had its instinct. It experienced a vague but menacing apprehension that subdued it and made it slink along at the man's heels, and that made it question eagerly every unusual movement of the man as if it expected to go seek shelter and build a fire. Fire was known by the dog, and it wanted fire, or else to burrow under the snow away from the air.

The frozen moisture of its breathing had settled on its fur in a fine powder of frost, and its jowls, muzzle, and eyelashes were especially whitened by its frozen breath. The man's red beard and moustache were likewise frosted, the deposit of ice and the form of ice and increase with every warm, moist breath he exhaled.

Continuing through the level stretch of woods for several miles, he dropped down a bank to the frozen bed of a small stream. This was Henderson Creek, and he knew that he was ten miles from the forks of the creek. Looking at his watch, he saw that it was ten o'clock. He was making four miles an hour, and his calculations told him his arrival at the forks would be at half-past twelve.
Directions: Using the following words and groups of words, write complete sentences.

1. (... ) (connect) (remote terminals) (telephone lines, leased lines, concentrators, etc.) to a front-end processor, which often permits access to one of several host processors at the data center.

2. During the last decades, (...) (consider) the density of storage on the tape and the speed with which we move tape, tape technology has become only a few hundred times better.

3. (...) (handle) more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day.

4. The data center manager who is concerned about (...) (reduce) operating costs or (...) (improve) performance of this central repository may introduce optical data disks.

5. (...) (at a higher system level) (current devices with optical data disks) (replace). The programmer must develop considerably more software to take advantage of optical data disks.

6. (...) (cold) (surprise), however.

7. He was a warm-whiskered man, but (...) (the hair on his face) (his high cheekbones and nose) (not) (protect).

8. The dog was subdued because of experiencing a vague but menacing apprehension that caused it to slink along at the man's heels, and that caused it to question eagerly every unusual movement of the man as if (...) (camp) (expect) (go) (seek shelter) and build a fire.

9. The man's red beard and moustache were likewise frosted, the deposit (...) (take) the form of ice and (...) (increase) with every warm, most breath he exhaled.

10. (...) (celebrate) (decide) (event) (he) by eating his lunch there.
TRUE FALSE QUESTIONS

Directions: Use the reading to answer the following questions. Circle T for True and F for False for each of the following questions.

T F 1. A data manager may introduce optical data disks only if he does not have to be concerned with cost reductions or improvements in performance of the central repository.

T F 2. Data center personnel have to mount manually large numbers of reels of tape per day since the system must process a larger quantity of tape faster.

T F 3. If optical disks replace current devices, a programmer must develop an increased amount of software.

T F 4. Tape technology has become a few hundred times better in the last thirty years.

T F 5. Access to one of several host processors at the data center is made possible by the connection of remote terminals to a front-end processor.

T F 6. The man's red beard and mustache were frosted because of his sweat.

T F 7. The man decided to eat his lunch at 12:30 at the forks of the creek.

T F 8. Since the man had a beard, his face was protected from the cold.

T F 9. The man expected the temperature to be at least 50° below zero.

T F 10. The dog watched the man carefully to see if he would go to a safe place to build a fire.
Common Language Version

Computer Science Text—Syntactically Altered Version

Fiction Text—Authentic Version
READING COMPREHENSION AND WRITING TEST

NAME__________________________________________

MAJOR__________________________________________

NATIVE LANGUAGE__________________________

AGE__________ CIRCLE ONE: FRESHMAN
SOPHOMORE
JUNIOR
SENIOR
GRADUATE STUDENT

STARTING TIME:__________

ENDING TIME:__________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
INSTRUCTIONS

Read the given texts carefully in order to answer the True/False questions afterwards. When you start reading the texts, record your beginning time on the cover sheet and when you have finished the texts, record the time again. Answer the True/False questions using the texts for reference.

Ten passages appear in italics within the texts. Within such a passage, words or groups of words are enclosed in parentheses ( ). The word or word groups are listed in alphabetic order. You are to make complete sentences of the passages in italics using the words in parentheses.

You may:
- change the order of the groups of words;
- change the forms of words (for example, adjective into adverb or verb into noun);
- conjugate the verb;
- add some words if you feel it is necessary to complete a sentence.

All the words or word groups within the ( ) have to be used.
The word order within the ( ) must remain the same.
For each ( . . ) you may add any number of words you would like or you may choose not to add any words.

If there are alternative solutions for any sentence, write the one you thought of first.

Write the sentences on the answer sheet.

EXAMPLE

( . . ) (computer simulations) (look) (use) the past and future of galaxies. The computers are programmed with data on the mass and size and relative locations of the galaxies.

POSSIBLE ANSWERS

1. Computer simulations are used to look into the past and future of galaxies.

2. Computer simulations can be used to look into the past and future of galaxies.

3. One can use computer simulations to look into the past and future of galaxies.
PART ONE

This text is concerned with computer science. It deals with the storage of great numbers of data in a computer.

MASS STORAGE SYSTEMS AND EVOLUTION OF DATA CENTER ARCHITECTURES (C)

The investment that an enterprise makes in equipment that processes data was once concentrated in the data center. The architecture of that data center is now changing into what might better be called a collection of cooperating subsystems that encompasses the geographic extent of the enterprise. (connect) (remote terminals) (telephone lines, leased lines, concentrators, etc.) to a front-end processor which often permits access to one of several host processors at the data center. Inside the data center, a single vendor may supply multiple hosts although processors representative of several vendors are becoming more common. If the data center itself does not contain multivendor hosts, the enterprise may well face a problem similar to having multiple data centers, located in different areas, that have grown separately around products from one particular vendor. Sometimes the several vendors are represented for reasons that are lost in history, but the fact remains that users throughout an enterprise desire access to files and other services that are distributed over several host processors. Associated with these data centers is the need for a central data repository of some description, whether it is based on paper, microfilm, or magnetic tape.

The typical central data repository in the DP world is the magnetic tape library. By "central repository" we mean the place from which data volumes are brought for entry into the system. The general paradigm is that one fetches the tape reel from the repository and mounts it on a drive. Then data are moved from the tape volumes to secondary storage (i.e., disks) for processing. Of course, once the reels are mounted, some data are processed directly from the magnetic tape because of the very short access time to the "next record." During the last decades, (consider) the density of storage of the tape and the speed with which we move tape, tape technology has become only a few hundred times better. During that same period of time, when we consider the number of instructions that the computer executes per second and the increased power of an individual instruction, the host computer (i.e., the central
processor and main memory) has increased in power several thousand
times. This disparity in rate of improvement places great pressures on the
system to "handle more tape faster."

\begin{align*}
&\text{\ldots} \text{(handle) more tape faster, many data center personnel must}
&\text{mount manually hundreds to thousands of tape reels per day. This is a}
&\text{labor-intensive, error-prone operation. The problems of erroneous tape}
&\text{mounts and the associated time delays lead many managers to keep their}
&\text{data bases "on line" on spinning magnetic disks (i.e., secondary storage).}
&\text{The cost of doing this includes the 100:1 (or more) differential in media}
&\text{costs plus the costs of relatively large amounts of electrical power and}
&\text{floor space that disk drives require.}
\end{align*}

\textit{The data center manager who is concerned about (\ldots) (reduce)
operating costs or (\ldots) (improve) performance of this central repository
may introduce optical data disks.} Such authors as Kenny and Brodie
have anticipated these disks for some time now. The efforts in the video
disk field of consumer electronics has stimulated R & D in the area of
optical data disks. By now, however, it is not possible to purchase
"off-the-shelf" devices for introduction into a data center.

Optical data disks promise that more bits can be stored on one
disk at lower costs per bit. However, they tend to employ a write-once,
read-many-times technology that does not permit that data are replaced
plug-compatible at the device level. (\ldots) (at a higher system level)
\textit{(current devices with optical data disks) (replace).} The programmer must
develop considerably more software to take advantage of optical data
disks. Even though much magnetic tape is used in a write-once,
read-many-times manner, a psychological barrier to the introduction of a
new technology that forces this mode of operation remains.
PART II
This text is from a short story about a man who is fighting against cold and the elements to survive.

From: TO BUILD A FIRE

Day had broken cold and gray, extremely cold and gray, when the man turned aside from the main Yukon River trail and climbed the high earth-bank, where a vague and little-travelled trail led eastward through the timberland. It was a steep bank, and he paused for breath at the top, excusing the act to himself by looking at his watch. It was nine o'clock.

The man flung a look back along the way he had come. The Yukon River lay a mile wide and hidden under three feet of ice. On top of this ice were as many feet of snow. It was all pure white, rolling in gentle waves where the ice-jams of the freeze-up had formed. North and south, as far as his eye could see, it was unbroken white.

As he turned to go on, he spat pensively. There was a sharp, explosive crackle that startled him. He spat again. And again, in the air, before it could fall to the snow, the spittle cracked. He knew that at fifty below zero spittle cracked in the air. Undoubtedly it was colder than fifty below--how much colder he did not know. But the temperature did not matter. He was bound for the old camp on the left fork of Henderson Creek, where the boys were already. They had come over from the Indian Creek country, while he had come the roundabout way to take a look at the possibilities of getting out logs from the islands in the Yukon in the spring. He would be in the camp by six o'clock; a little after dark, it was true, but the boys would be there, a fire would be going, and a hot supper would be ready. As for lunch, he pressed his hand against the protruding bundle under his jacket. It was also under his shirt, wrapped up in a handkerchief and lying against the naked skin. It was the only way to keep the biscuits from freezing. He smiled agreeably to himself as he thought of these biscuits, each cut open and sopped in bacon grease and each enclosing a generous slice of fried bacon. In fact, he carried nothing but the lunch wrapped in the handkerchief. (. . .) (cold) (surprise), however. It certainly was cold, he concluded, as he rubbed his numb nose and cheekbones with his mittened hand. He was a
warm-whiskered man, but (. . .) (his high cheekbones and nose) (not) (protect) (the hair on his face).

At the man's heels trotted a dog, a big native wolf-dog, gray-coated and without any visible or temperamental difference from its brother, the wild wolf. The animal was depressed by the tremendous cold. It knew it was no time for travelling. It was seventy-five below zero. The freezing-point is thirty-two above zero. The dog did not know anything about thermometers. But the brute had its instinct. It experienced a vague but menacing apprehension that subdued it and made it slink along at the man's heels, and that made it question eagerly every unusual movement of the man as if (. . .) (camp) (expect) (go) (seek shelter) and build a fire. The dog knew fire, and it wanted fire, or else to burrow under the snow away from the air.

The frozen moisture of its breathing had settled on its fur in a fine powder of frost, and its jowls, muzzle, and eyelashes were especially whitened by its frozen breath. The man's red beard and moustache were likewise frosted, the deposit (. . .) (take) the form of ice and (. . .) (increase) with every warm, moist breath he exhaled.

He continued through the level stretch of woods for several miles and dropped down a bank to the frozen bed of a small stream. This was Henderson Creek, and he knew that he was ten miles from the forks of the creek. He looked at his watch. It was ten o'clock. He was making four miles an hour, and he calculated that he would arrive at the forks at half-past twelve. (. . .) (celebrate) (decide) (event) (he) by eating his lunch there.
1. (connect) (remote terminals) (telephone lines, leased lines, concentrators, etc.) to a front-end processor, which often permits access to one of several host processors at the data center.

2. During the last decades, (consider) the density of storage on the tape and the speed with which we move tape, tape technology has become only a few hundred times better.

3. (handle) more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day.

4. The data center manager who is concerned about (reduce) operating costs or (improve) performance of this central repository may introduce optical data disks.

5. (at a higher system level) (current devices with optical data disks) (replace). The programmer must develop considerably more software to take advantage of optical data disks.

6. (cold) (surprise), however.

7. He was a warm-whiskered man, but (the hair on his face) (his high cheekbones and nose) (not) (protect).

8. The dog was subdued because of experiencing a vague but menacing apprehension that caused it to slink along at the man's heels, and that caused it to question eagerly every unusual movement of the man as if (camp) (expect) (go) (seek shelter) and build a fire.

9. The man's red beard and moustache were likewise frosted, the deposit (take) the form of ice and (increase) with every warm, most breath he exhaled.

10. (celebrate) (decide) (event) (he) by eating his lunch there.
TRUE FALSE QUESTIONS

Directions: Use the reading to answer the following questions. Circle T for True and F for False for each of the following questions.

T  F  1. A data manager may introduce optical data disks only if he does not have to be concerned with cost reductions or improvements in performance of the central repository.

T  F  2. Data center personnel have to mount manually large numbers of reels of tape per day since the system must process a larger quantity of tape faster.

T  F  3. If optical disks replace current devices, a programmer must develop an increased amount of software.

T  F  4. Tape technology has become a few hundred times better in the last thirty years.

T  F  5. Access to one of several host processors at the data center is made possible by the connection of remote terminals to a front-end processor.

T  F  6. The man's red beard and mustache were frosted because of his sweat.

T  F  7. The man decided to eat his lunch at 12:30 at the forks of the creek.

T  F  8. Since the man had a beard, his face was protected from the cold.

T  F  9. The man expected the temperature to be at least 50° below zero.

T  F  10. The dog watched the man carefully to see if he would go to a safe place to build a fire.

132
APPENDIX B

TEST BOOKLETS FOR
THE PARAGRAPH EXPERIMENT
WRITING SAMPLE

Directions: Write a logical, well-formed paragraph on the subject of Gravity containing five (5) sentences, using every word on the list below. You may change the form of any of the given words, but you must use all of the information given.

Use these words:
Isaac Newton falling objects force of attraction moon earth tide theory
WRITING SAMPLE

Directions: Write a logical, well-formed paragraph on the subject of Newspapers containing five (5) sentences, using every word on the list below. You may change the form of any of the given words, but you must use all of the information given.

Use these words:
- information
- business
- effect
- feature
- advertisements
- news of the day
- reporters
APPENDIX C

Test Booklets

Reading ST vs. CL Syntax

Within a Computer Science Topic
READING COMPREHENSION TEST

NAME ________________________________

MAJOR ______________________________

NATIVE LANGUAGE ______________________

AGE _______ CIRCLE ONE: FRESHMAN
SOPHOMORE
JUNIOR
SENIOR
GRADUATE STUDENT

STARTING TIME: __________

ENDING TIME: __________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
MASS STORAGE SYSTEMS AND
EVOLUTION OF DATA CENTER ARCHITECTURES

The investment that an enterprise makes in data processing equipment was once concentrated in the data center. The architecture of that data center is now changing into what might better be called a collection of cooperating subsystems that encompasses the geographic extent of the enterprise. Via telephone lines, leased lines, concentrators, etc., remote terminals are connected to a front-end processor, which often permits access to one of several host processors at the data center. Inside the data center, the multiple hosts may be supplied by a single vendor, although processors representative of several vendors are becoming more common. If the data center itself does not contain multivendor hosts, the enterprise may well face a problem similar to having multiple data centers, located in different areas, that have grown separately around products from a particular vendor. Sometimes the several vendors are represented for reasons that are lost in history, but the fact remains that users throughout an enterprise desire access to files and other services that are distributed over several host processors. Associated with these data centers is the need for a central data repository of some description, whether it is based in paper, microfilm, or magnetic tape.

The typical central data repository in the DP world is the magnetic tape library. By "central repository" we mean the place from which data volumes are brought for entry into the system. The general paradigm is that the tape reel is fetched from the repository and mounted on a drive. Then data are moved from the tape volumes to secondary storage (i.e., disks) for processing. Of course, once the reels are mounted some data are processed directly from magnetic tape because of the very short access time to the "next record." During the last decades, considering the density of storage on the tape and the speed with which we move tape, tape technology has become only a few hundred times better. During that same period of time, considering the instructions executed per second and the increased power of an individual instruction, the host computer (i.e., the central processor and main memory) has increased in power several thousand times. This disparity in rate of improvement places great pressures on the system to "handle more tape faster."

In handling more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day. This is a labor-intensive, error-prone operation. The problems of erroneous tape
mounts and the associated time delays lead many managers to keep their data bases "on-line" on spinning magnetic disks (i.e., secondary storage). The cost of doing this includes the 100:1 (or more) differential in media costs plus the costs of the relatively large amounts of electrical power and floor space that are required by disk drives.

New Technology

The data center manager who is concerned about reducing operating costs or improving performance of this central repository may introduce optical data disks. These disks have been anticipated for some time now by such authors as Kenny and Brodie. R & D in the area of optical data disks has been stimulated by the efforts in the video disk field of consumer electronics. At this time, however, it is not possible to purchase "off-the-shelf" devices for introduction into a data center.

Optical data disks promise higher storage density and lower per-bit costs but tend to employ a write-once, read-many-times technology that does not lend itself to plug-compatible replacement at the device level. Replacement at a higher system level is required; this implies considerable software development to take advantage of these promising device characteristics. Even though much magnetic tape is used in a write-once, read-many-times manner, there remains a psychological barrier to the introduction of a new technology that forces this mode of operation.

An interesting speculative investigation by George P. Copeland, couples these characteristics with the probability of judicious use of surrogates and time stamps in a nondeletion policy, improved data base designs are possible. Although most people tend to think of optical data disks in terms of archival storage, the thinking exhibited by Copeland in this article could blur the distinction between secondary and archival storage.

New Architectures

Whether optical or magnetic media are used, the accumulation of massive amounts of data in "one place" highlights the necessity to control the access to and the movement of data to and from that "place." In fact, the experience of those who have introduced mass storage system (i.e., mechanical replacements for the magnetic tape library) has shown that the greatest problem was to get the architecture straight so that the library was accessible to all who needed it. Exactly what is meant by "getting
the architecture straight" may be debatable; however, it includes staging
the data to secondary storage where it can be used by an application
program in a manner that does not require recompilation of the
application programs in the library. It also includes the development of
common indices and directories to permit the application program to
locate and retrieve the desired file from a library serving multiple hosts.

Some members of the MSS community have requested "clean
interfaces" to allow the economical introduction of new devices. Certainly
the I/O structure of the present-day computer does not constitute a clean
interface, primarily because the operating systems assume knowledge of
the physical characteristics of the device. Clean interfaces come from a
functional model of the architecture that is quite modular. The functions of
the modules and the description of what passes between them is readily
understood by user and vendor alike. The MSS community would benefit
from having a simple functional model upon which to focus its attention
rather than focusing primarily on physical characteristics.

Fortunately, operating systems have been evolving from
monolithic giants toward small kernalized systems. Many of the functions
(i.e., file management, user interface, and I/O drivers) are removed from
the monolithic predecessors and become modular application programs
with clean interfaces to the smaller kernalized operating systems. Any
one of the modules can be independently improved.

The International Standards Organization has provided a
functional model in the data communications field, the open systems
interconnect reference model. The ISO/OSI reference model incorporates
seven layers or modules for which the functions but not the
implementations are defined. As layered modules, the higher layers
depend upon the services of the lower layers. The interfaces between the
layers are defined in terms of the signals and services offered across the
interface to the adjacent layer. These constitute "clean interfaces"
because a brand-new implementation may be introduced to replace a
single or small set of layers (modules) without having to redesign (or
emulate) the entire set.

Staging is a generic term for the movement of data into and out of
"the system." The connectivity permitted by the I/O architecture of most
computer systems is very restrictive. For example, most host computers
do not permit the direct transfer of a data file from magnetic tape to
magnetic disk; they require reading the tape into main memory and then
writing it out to the disk. The term "backfill staging" has been used to
indicate the possibility of the host directing such a transfer without the movement of data into and out of the main memory. Of course, this requires some type of control capability outside of the host to monitor the transfer as well as the physical path.

Application Studies

Whenever data are moved across any boundary, there must be a body of software running on some piece of hardware to supervise that transfer and ensure its success or report its failure. This software/hardware combination may be the portion of the operating system devoted to I/O, a communications protocol (monolithic or layered), or it may be an outboard controller devoted to the control of the library. Bernard T. O'Lear and Joseph H. Cloy show that the requisite software need not all be on the same machine. Understanding this requisite software and deciding where it will operate is an integral step in "getting the architecture straight."

Staging, or the movement of all of the data contained in a physical volume into and out of the system, implies a preplanned movement. The planning is generally included in the application program. The user mounts a tape reel or disk volume and the application program running in the host computer selects the desired file from the volume. This occurs whether the retrieval is manual or uses some mechanical robot under system control. When such movement is possible under system control, the problem becomes one of separating and clustering sets of files independently of the physical volumes and providing a mechanism to have desired files more readily available. This is frequently termed "migration" and implies a certain randomness beyond user control.

The architectures of information processing systems have been evolving toward a high-speed local network permitting ready communication among multiple hosts, multiple peripheral controllers, and concentrator/manager processors for terminals and/or connections to other systems. This trend is driven by the need for improved communication among data-hungry monster hosts and their central repositories.

The data communication community is being aided by the existence of the ISO/OSI reference model. Comparisons of protocol architecture of widely different communication networks can be performed by comparing the functions provided to the functions in the individual layers of the OSI reference model. It provides a common point of
understanding between users and vendors.

The mass storage community could be aided by developing a common reference model that builds upon and extends the OSI reference model to include the services needed for mass storage systems serving multiple hosts.
TRUE/FALSE QUESTIONS

T  F  1. If optical disks replace current devices, a programmer must develop an increased amount of software.

T  F  2. Data center personnel have to manually mount large numbers of reels of tape per day since the system must process a larger quantity of tape faster.

T  F  3. If the user wants to replace a single or small set of layers, he has to design a whole new set.

T  F  4. Tape technology has become a few hundred times better in the last thirty years.

T  F  5. Since the operating systems assume that the user knows the physical characteristics of the device, the computer's I/O structure is a clean interface.

T  F  6. Telephone lines connect remote terminals to a front-end processor which may permit access to a host processor at the data center.

T  F  7. It is impossible for the system to separate and cluster sets of files under system control.

T  F  8. A data manager may introduce optical data disks only if he does not have to reduce costs or improve performance of the central repository.

T  F  9. Backfill staging means that the host has to move data in and out of the main memory while it directs the transfer of a data file.

T  F  10. Getting the architecture straight means that application programs in the library are automatically recompiled when data is staged to secondary storage.
Common Language Version

Syntactically Altered Text
READING COMPREHENSION TEST

NAME______________________________

MAJOR______________________________

NATIVE LANGUAGE_____________________

AGE__________ CIRCLE ONE: FRESHMAN

SOPHOMORE

JUNIOR

SENIOR

GRADUATE STUDENT

STARTING TIME:__________

ENDING TIME:__________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
The investment that an enterprise makes in data processing equipment was once concentrated in the data center. The architecture of that data center is now changing into what might better be called a collection of cooperating subsystems that encompasses the geographic extent of the enterprise. Telephone lines, leased lines, concentrators, etc., connect remote terminals to a front-end processor. This processor often permits access to one of several host processors at the data center. Inside the data center, the multiple hosts may be supplied by a single vendor, although processors representative of several vendors are becoming more common. If the data center itself does not contain multivendor hosts, the enterprise may well face a problem similar to having multiple data centers, located in different areas, that have grown separately around products from a particular vendor. Sometimes the several vendors are represented for reasons that are lost in history, but the fact remains that users throughout an enterprise desire access to files and other services that are distributed over several host processors. Associated with these data centers is the need for a central data repository of some description, whether it is based in paper, microfilm, or magnetic tape.

The typical central data repository in the DP world is the magnetic tape library. By "central repository" we mean the place from which data volumes are brought for entry into the system. The general paradigm is that the tape reel is fetched from the repository and mounted on a drive. Then data are moved from the tape volumes to secondary storage (i.e., disks) for processing. Of course, once the reels are mounted some data are processed directly from magnetic tape because of the very short access time to the "next record." During the last decades, if we consider that the tape stores more data in a smaller space and moves much faster, tape technology has become only a few hundred times better. During that same period of time, considering the instructions executed per second and the increased power of an individual instruction, the host computer (i.e., the central processor and main memory) has increased in power several thousand times. This disparity in rate of improvement places great pressures on the system to "handle more tape faster."
Because the system must handle more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day. This is a labor-intensive, error-prone operation. The problems of erroneous tape mounts and the associated time delays lead many managers to keep their data bases "on-line" on spinning magnetic disks (i.e., secondary storage). The cost of doing this includes the 100:1 (or more) differential in media costs plus the costs of the relatively large amounts of electrical power and floor space that are required by disk drives.

**New Technology**

A data center manager may introduce optical data disks if he must reduce operating costs or improve performance of this central repository. These disks have been anticipated for some time now by such authors as Kenny and Brodie. R & D in the area of optical data disks has been stimulated by the efforts in the video disk field of consumer electronics. At this time, however, it is not possible to purchase "off-the-shelf" devices for introduction into a data center.

Optical data disks promise higher storage density and lower per-bit costs but tend to employ a write-once, read-many-times technology that does not lend itself to plug-compatible replacement at the device level. The user must replace current devices with optical data disks at a higher system level. The programmer must develop considerably more software to take advantage of optical data disks. Even though much magnetic tape is used in a write-once, read-many-times manner, there remains a psychological barrier to the introduction of a new technology that forces this mode of operation.

**New Architectures**

Whether optical or magnetic media are used, the accumulation of massive amounts of data in "one place" highlights the necessity to control the access to and the movement of data to and from that "place." In fact, the experience of those who have introduced mass storage system (i.e., mechanical replacements for the magnetic tape library) has shown that the greatest problem was to get the architecture straight so that the library was accessible to all who needed it. The meaning of "getting the architecture straight" may be debatable; however, the act of getting the architecture straight means that the system stages the data to secondary storage where an application program can use it so that application programs in the library do not have to be recompiled. It also includes the
development of common indices and directories to permit the application program to locate and retrieve the desired file from a library serving multiple hosts.

Some members of the MSS community have requested "clean interfaces" to allow the economical introduction of new devices. Certainly the I/O structure of the present-day computer does not constitute a clean interface, primarily because the operating systems assume that the user knows the physical characteristics of the device. Clean interfaces come from a functional model of the architecture that is quite modular. The functions of the modules and the description of what passes between them is readily understood by user and vendor alike. The MSS community would benefit from having a simple functional model upon which to focus its attention rather than focusing primarily on physical characteristics.

Fortunately, operating systems have been evolving from monolithic giants toward small kernalized systems. Many of the functions (i.e., file management, user interface, and I/O drivers) are removed from the monolithic predecessors and become modular application programs with clean interfaces to the smaller kernalized operating systems. Any one of the modules can be independently improved.

The International Standards Organization has provided a functional model in the data communications field, the open systems interconnect reference model. The ISO/OSI reference model incorporates seven layers or modules for which the functions but not the implementations are defined. As layered modules, the higher layers depend upon the services of the lower layers. The interfaces between the layers are defined in terms of the signals and services offered across the interface to the adjacent layer. These constitute "clean interfaces" because the user may implement something new to replace a single or small set of layers (modules) without having to redesign (or emulate) the entire set.

Staging is a generic term for the movement of data into and out of "the system." The connectivity permitted by the I/O architecture of most computer systems is very restrictive. For example, most host computers do not permit the direct transfer of a data file from magnetic tape to magnetic disk; they require reading the tape into main memory and then writing it out to the disk. The term "backfill staging" indicates that the host can direct such a transfer without moving data into and out of the main memory. Of course, this requires some type of control capability outside
of the host to monitor the transfer as well as the physical path.

Application Studies

Whenever data are moved across any boundary, there must be a body of software running on some piece of hardware to supervise that transfer and ensure its success or report its failure. This software/hardware combination may be the portion of the operating system devoted to I/O, a communications protocol (monolithic or layered), or it may be an outboard controller devoted to the control of the library. Bernard T. O'Lear and Joseph H. Cloy show that the requisite software need not all be on the same machine. Understanding this requisite software and deciding where it will operate is an integral step in "getting the architecture straight."

Staging, or the movement of all of the data contained in a physical volume into and out of the system, implies a preplanned movement. The planning is generally included in the application program. The user mounts a tape reel or disk volume and the application program running in the host computer selects the desired file from the volume. This occurs whether the retrieval is manual or uses some mechanical robot under system control. When such movement is possible under system control, the problem is, can the system separate and cluster sets of files independently of the physical volumes and provide a mechanism to have desired files more readily available. This is frequently termed "migration" and implies a certain randomness beyond user control.

The architectures of information processing systems have been evolving toward a high-speed local network permitting ready communication among multiple hosts, multiple peripheral controllers, and concentrator/manager processors for terminals and/or connections to other systems. This trend is driven by the need for improved communication among data-hungry monster hosts and their central repositories.

The data communication community is being aided by the existence of the ISO/OSI reference model. Comparisons of protocol architecture of widely different communication networks can be performed by comparing the functions provided to the functions in the individual layers of the OSI reference model. It provides a common point of understanding between users and vendors.

The mass storage community could be aided by developing a
common reference model that builds upon and extends the OSI reference model to include the services needed for mass storage systems serving multiple hosts.
Directions: Use the reading to answer the following questions.
Circle T for true and F for false for each of the following questions.

TRUE/FALSE QUESTIONS

T  F  1. Since the operating systems assume that the user knows the physical characteristics of the device, the computer's I/O structure is a clean interface.

T  F  2. Telephone lines connect remote terminals to a front-end processor which may permit access to a host processor at the data center.

T  F  3. Backfill staging means that the host has to move data in and out of the main memory while it directs the transfer of a data file.

T  F  4. A data manager may introduce optical data disks only if he does not have to reduce costs or improve performance of the central repository.

T  F  5. If optical disks replace current devices, a programmer must develop an increased amount of software.

T  F  6. Data center personnel have to manually mount large numbers of reels of tape per day since the system must process a larger quantity of tape faster.

T  F  7. Getting the architecture straight means that application programs in the library are automatically recompiled when data is staged to secondary storage.

T  F  8. If the user wants to replace a single or small set of layers, he has to design a whole new set.

T  F  9. Tape technology has become a few hundred times better in the last thirty years.

T  F  10. It is impossible for the system to separate and cluster sets of files under system control.
Common Language Version

Authentic Text
APPENDIX D

Test Booklets

READING FICTION TEXT
READING COMPREHENSION TEST

NAME_____________________________________

MAJOR_____________________________________

NATIVE LANGUAGE___________________________

AGE_________ CIRCLE ONE: FRESHMAN
SOPHOMORE
JUNIOR
SENIOR
GRADUATE STUDENT

STARTING TIME:_________

ENDING TIME:_________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
This text is from a short story about a man who is fighting against cold and the elements to survive.

TO BUILD A FIRE

Day had broken cold and gray, extremely cold and gray, when the man turned aside from the main Yukon River trail and climbed the high earth-bank, where a vague and little-traveled trail led eastward through the timberland. It was a steep bank, and he paused for breath at the top, excusing the act to himself by looking at his watch. It was nine o'clock.

The man flung a look back along the way he had come. The Yukon River lay a mile wide and hidden under three feet of ice. On top of this ice were as many feet of snow. It was all pure white, rolling in gentle waves where the ice-jams of the freeze-up had formed. North and south, as far as his eye could see, it was unbroken white.

As he turned to go on, he spat pensively. There was a sharp, explosive crackle that startled him. He spat again. And again, in the air, before it could fall to the snow, the spittle crackled. He knew that at fifty below zero spittle crackled in the air. Undoubtedly it was colder than fifty below--how much colder he did not know. But the temperature did not matter. He was bound for the old camp on the left fork of Henderson Creek, where the boys were already. They had come over from the Indian Creek country, while he had come the roundabout way to take a look at whether it would be possible to get out logs from the islands in the Yukon in the spring. He would be in the camp by six o'clock; a little after dark, it was true, but the boys would be there, a fire would be going, and a hot supper would be ready. As for lunch, he pressed his hand against the protruding bundle under his jacket. It was also under his shirt, wrapped up in a handkerchief and lying against the naked skin. It was the only way to keep the biscuits from freezing. He smiled agreeably to himself as he thought of those biscuits, each cut open and sopped in bacon grease and each enclosing a generous slice of fried bacon.

He plunged in among the big spruce trees. The trail was faint. A foot of snow had fallen since the last sled had passed over, and he was glad he was without a sled, traveling light. In fact, he carried nothing but the lunch wrapped in the handkerchief. The cold surprised him, however. It certainly was cold, he concluded, as he rubbed his numb nose and cheek-bones with his mitten hungry hand. He was a warm-whiskered man, but the hair on his face did not protect the high cheekbones and the eager nose that thrust itself aggressively into the frosty air.
At the man's heels trotted a dog, a big native wolf-dog, gray-coated and without any visible or temperamental difference from its brother, the wild wolf. The tremendous cold depressed the animal. It knew that it was no time for traveling. It was seventy-five below zero. The freezing-point is thirty-two above zero. The dog did not know anything about thermometers. But the brute had its instinct. It experienced a vague but menacing apprehension that subdued it and made it slink along at the man's heels, and that made it question eagerly every unusual movement of the man as if it expected him to go to camp and build a fire. The dog knew fire, and it wanted fire, or else to burrow under the snow away from the air.

The frozen moisture of its breathing had settled on its fur in a fine powder of frost, and its jowls, muzzle, and eyelashes were especially whitened by its frozen breath. The man's red beard and mustache were likewise frosted, the deposit taking the form of ice and increasing with every warm, moist breath he exhaled.

He continued through the level stretch of woods for several miles and dropped down a bank to the frozen bed of a small stream. This was Henderson Creek, and he knew that he was ten miles from the forks of the creek. He looked at his watch. It was ten o'clock. He was making four miles an hour, and he calculated that he would arrive at the forks at half-past twelve. He decided he would celebrate the event by eating his lunch there.
TRUE/FALSE QUESTIONS

Directions: Use the reading to decide whether the following statements are true or false. Circle T for True and F for False for each question.

1. When the man looked back at where he had already travelled, he saw three feet of ice covered by three feet of snow.  

2. The man's red beard and mustache were frosted because of his sweat.  

3. At the camp, the boys would have already built a fire and would have prepared supper by the time the man arrived.  

4. The man decided to eat his lunch at 12:30 at the forks of the creek.  

5. The man kept his lunch under his jacket so he would have his hands free to build a fire.  

6. The man's lunch was some biscuits with bacon grease and bacon inside them.  

7. Since the man had a beard, his face was protected from the cold.  

8. The man followed a large wolf-dog on the trip.  

9. The man expected the temperature to be at least 50° below zero.  

10. The dog watched the man carefully to see if he would go to a safe place to build a fire.
Science/Technology Version

Syntactically Adapted
READING COMPREHENSION TEST

NAME________________________________________

MAJOR________________________________________

NATIVE LANGUAGE______________________________

AGE___________ CIRCLE ONE: FRESHMAN
SOHOMORE
JUNIOR
SENIOR
GRADUATE STUDENT

STARTING TIME: __________

ENDING TIME: __________

DIRECTIONS:

Read the text carefully in order to answer True/False questions afterwards. Record the time you begin to read and the time you complete the task.
This text is from a short story about a man who is fighting against cold and the elements to survive.

**TO BUILD A FIRE**

Day had broken cold and gray, extremely cold and gray, when the man decided to turn aside from the main Yukon River trail and to climb the high earth-bank, where a vague and little-traveled trail led eastward through the timberland. The steepness of the bank caused him to pause for breath at the top, excusing the act to himself by looking at his watch. It was nine o'clock.

The man's look was flung back along the way he had come. The Yukon River lay a mile wide and was hidden under three feet of ice. This ice was covered by as many feet of snow. It was all pure white, rolling in gentle waves where the ice-jams of the freeze-up had formed. North and south, as far as could be seen by his eyes, it was unbroken white.

As he turned to go on, he spat pensively. He was startled by a sharp, explosive crackle. He spat again. And again, in the air, before it could fall to the snow, the spittle crackled. He knew that spittle was made to crack in the air at fifty below zero. Undoubtedly it was colder than fifty below--how much colder was not known. But the temperature did not matter. He was bound for the old camp on the left fork of Henderson Creek, where the boys were already. They had come over from the Indian Creek country, while he had come the roundabout way to take a look at the possibilities of getting out logs from the islands in the Yukon in the spring. He would be in the camp by six o'clock; a little after dark, it was true, but the boys would be there, a fire would have been built and a hot supper would have been prepared. As for lunch, he pressed his hand against the protruding bundle under his jacket. It was also under his shirt, wrapped up in a handkerchief and lying against the naked skin. It was the only way to keep the biscuits from freezing. He smiled agreeably to himself as he thought of those biscuits, each had been cut open and sopped in bacon grease and a generous slice of fried bacon was enclosed in each.

He plunged in among the big spruce trees. The trail was faint. A foot of snow had fallen since the last sled had passed over, and he was glad he was without a sled, traveling light. In fact, nothing was carried by him but the lunch wrapped in the handkerchief. He was surprised, however, at the cold. It certainly was cold, he concluded, as he rubbed his numb nose and cheek-bones with his mitten hand. He was a
warm-whiskered man, but his high cheekbones and the eager nose that thrust itself aggressively into the frosty air were not protected by the hair on his face.

The man was followed by a dog trotting at his heels, a big native wolf-dog, gray-coated and without any visible or temperamental difference from its brother, the wild wolf. The animal was depressed by the tremendous cold. Its knowledge told it that it was no time for traveling. It was seventy-five below zero. The freezing-point is thirty-two above zero. Nothing about thermometers was know by the dog. But the brute had its instinct. The dog was subdued because of experiencing a vague but menacing apprehension that caused it to slink along at the man’s heels and that caused it to question eagerly every unusual movement of the man as if the man were expected to go to camp and to build a fire. Fire was known by the dog, and it wanted fire, or else to burrow under the snow away from the air.

The frozen moisture of its breathing had settled on its fur in a fine powder of frost, and its jowls, muzzle, and eyelashes were especially whitened by its frozen breath. The man’s red beard and mustache were likewise frosted by ice. This frost was increased with every warm, moist breath which was exhaled by the man.

He continued through the level stretch of woods for several miles and dropped down a bank to the frozen bed of a small stream. This was Henderson Creek, and he knew that he was ten miles from the forks of the creek. He looked at his watch. It was ten o’clock. He was making four miles an hour, and he calculated that he would arrive at the forks at half-past twelve. He decided to have a celebration of the event by eating his lunch there.
TRUE/FALSE QUESTIONS

Directions: Use the reading to decide whether the following statements are true or false. Circle T for True and F for False for each question.

T F 1. When the man looked back at where he had already travelled, he saw three feet of ice covered by three feet of snow.

T F 2. The man's red beard and mustache were frosted because of his sweat.

T F 3. At the camp, the boys would have already built a fire and would have prepared supper by the time the man arrived.

T F 4. The man decided to eat his lunch at 12:30 at the forks of the creek.

T F 5. The man kept his lunch under his jacket so he would have his hands free to build a fire.

T F 6. The man's lunch was some biscuits with bacon grease and bacon inside them.

T F 7. Since the man had a beard, his face was protected from the cold.

T F 8. The man followed a large wolf-dog on the trip.

T F 9. The man expected the temperature to be at least 50° below zero.

T F 10. The dog watched the man carefully to see if he would go to a safe place to build a fire.
REFERENCES
REFERENCES


Healy, A. (1981). The effects of visual similarity on proofreading for


Tannenbaum, P. H., & Williams, F. (1968). Generation of active and passive sentences as a function of subject or object focus. *Journal of
Verbal Learning and Verbal Behavior, 7, 246-250.


SAMENVATTING

Schrijven en lezen worden beschouwd als twee onmisbare delen van een schriftelijke communicatie-keten. Schrijvers moeten een tekst schrijven die kan worden begrepen door specifieke lezers en omgekeerd, lezers moeten in staat zijn te begrijpen wat schrijvers produceren. Syntaxis is zeker een factor die bepaalt hoe teksten worden geschreven en met welk gemak zij worden gelezen.

In deze dissertatie worden enige vragen onderzocht die verband houden met het effect van syntaxis op zowel lezen als schrijven en dan specifiek binnen het WT (wetenschappelijk- en technisch) register.

Is er een verschil in de wijze waarop wetenschappelijke en technische teksten worden geschreven in vergelijking met teksten over een niet-WT-onderwerp?

Als er een verschil bestaat, is dat dan een functie van de schrijver, bijv. cognitief systeem, achtergrondkennis van het onderwerp waarover wordt geschreven (deskundige versus leek m.b.t. het onderwerp van de tekst), of talenkennis (moedertaal of eerste taal (T1) versus tweede of vreemde taal (T2))? Is het een functie van de tekst zelf - het onderwerp waarover wordt geschreven? Is het een functie van de lezer of groep lezers van de tekst - specialisten op het gebied van het onderwerp versus leken? Of is het verschil het resultaat van een interactie tussen kennis en vaardigheden van de schrijver en de kenmerken van de te produceren tekst.

Is er een verschil in de wijze waarop WT-teksten worden gelezen, vergeleken met teksten in de omgangstaal (OT)? Als er een verschil bestaat is, dat dan een functie van de lezer - bijv. cognitief systeem, achtergrondkennis van het te lezen onderwerp, talenkennis - of is het een functie van de tekst - bijv. syntaxis, lexicon, discourse-structuur? Of is het verschil het resultaat van een interactie tussen kennis en vaardigheden van de lezer en
de kenmerken van de te lezen tekst? Deze dissertatie geeft althans beperkte antwoorden op bovenstaande vragen. Ook wordt het verband onderzocht tussen het lees- en schrijfproces in hun relatie tot WT-teksten.

In de schrijfexperimenten gingen de primaire vragen over wat een bepaalde schrijver er toe brengt bepaalde syntactische structuren te kiezen binnen een bepaald onderwerp of semantische context. De algemene conclusie die op grond van de resultaten getrokken kan worden is dat alle proefpersonen - T1 en T2 - deskundige en leek - de voorkeur geven aan het gebruik van WT-syntaxis in wetenschappelijke teksten en OT-syntaxis in niet-wetenschappelijke teksten.

Het probleem van de invloed van cotext (de verzameling van alle zinnen die een bepaalde zin in een tekst omringen) op de keuze van de syntaxis, werd geëlimineerd. Voor de meeste proefpersonen uit deze experimenten blijkt heel duidelijk dat schrijvers voor syntactische structurering een gelijkwaardige keus hebben tussen WT- en OT-register. Toch is het onderwerp of de semantische context zelf de primaire factor waardoor de keuze van de syntaxis door een schrijver wordt bepaald, ongeacht zijn of haar taalachtergrond of deskundigheid m.b.t. het onderwerp waarover wordt geschreven. Daarom is, althans voor het syntactische element van het WT-register, dat register een psycholinguïstische realiteit.

Voor het WT-register heeft het onderwerp van de tekst een sterkere invloed op het gebruik van WT-syntaxis dan de achtergrondkennis of de talenkennis. Het onderwerp van de tekst suggereert automatisch de te gebruiken syntaxis, ongeacht de achtergrond- en talenkennis van de schrijver. Schrijvers neigen ertoe hun linguïstische register op een psycholinguïstische wijze af te stemmen op het onderwerp; derhalve is het WT-register niet louter een zaak van linguïstische beschrijving.

Aangezien schrijvers erop zijn gericht hun linguïstische register af te stemmen op het onderwerp van de tekst, onafhankelijk van achtergrondkennis (deskundige vs leek) en talenkennis (T1 en T2), is technisch schrijven, zowel in de moedertaal als in een vreemde taal, niet alleen een zaak
van taal en van opleiding in specialistische kennis, maar ook van het verwerven van inzicht in psycholinguïstische aspecten van het schrijven.

Voor lezen was de primaire vraag of het WT-register met zijn veronderstelde moeilijkere structuren, zoals passiva, nominalisaties, deelwoorden en infinitieven, moeilijker is te begrijpen dan het OT-register.

Uit de leesexperimenten in deze dissertatie kan duidelijk worden geconcludeerd dat de syntaxis van het WT-register niet meer begripsmoeilijkheden veroorzaakt dan het OT-register voor alle groepen die aan het onderzoek hebben deelgenomen. Bovendien schijnen zowel T1, als T2 lezers niet erg gehinderd te worden bij het begrijpen van een tekst, zelfs als de syntaxis niet in tact is (zinselementen worden aangeboden in een alfabetische in plaats van een natuurlijke volgorde). Dat een afwezige syntactische structuur zowel door lezers in de moedertaal als in een vreemde taal in hoge mate kan worden begrepen, is een extra bewijs dat een conceptueel gestuurde en partiël parallelle strategie bij het verwerken van teksten geen volledige syntactische ontleiding nodig heeft. In het laatste hoofdstuk worden de praktische implicaties behandeld voor technische lezers en schrijvers en voor docenten technisch lezen en schrijven, zowel voor studenten in de moedertaal als in een vreemde taal.
For the ST (Science and Technology) register, the topic of the text has a stronger impact on ST syntax use than background knowledge (expert vs. nonexpert) or language knowledge (L1 vs L2).

Reading to acquire background knowledge of the field to be written about may be the most critical task for improving technical writing.

As long as there is some base level knowledge of the target language, it is more important to have background knowledge of the subject to be written in than to have enhanced proficiency in the target language for subject-specific topics.

Readability formulas should not be used as guidelines for technical writers (e.g., Duffy, 1985; Felkner, Redish, & Peterson, 1985; Selzer, 1983).

Since experts process information differently than nonexperts, the writer should take into account the expertise of the reader so that he or she can design the text to facilitate processing. (Huckin 1983; Samuels, 1988).

To date, human beings have been the only creatures to learn human language (Glass & Holyoak, 1986).

Money spent in space is well spent on earth.

With our increasing awareness of the interlocking nature of the world's ecosystems, the rapid destruction of the Brazilian rain forest is of vital concern to all nations. (Brown, et al, 1989).
In the face of rising world population and dwindling resources, western style materialism will remain an unfulfilled dream of most emerging nations. (Hartmann, et al, 1984).

Perhaps the best management system is a combination of Japanese bottom up and the European and American top down systems.
