The effect of syntactic simplification on reading EST texts as L1 and L2
Jan M. Ulijn and Judith B. Strother
Eindhoven University of Technology, The Netherlands and Florida Institute of Technology, U.S.A.

ABSTRACT

There has been a considerable suggestion in the literature that syntactic adaptation of a text would simplify it. This empirical study determined that the complexity of the syntax does not significantly affect the level of reading comprehension. While a complete conceptual and lexical analysis may be necessary for reading comprehension, a thorough syntactic analysis is not. This should be true for both expert and novice readers in a particular professional field, herein computer science.

Using American (L1 or E1, English as a native language) and Dutch (L2 or E2, English as a second, nonnative language) computer science and humanities majors reading in EST (English for Science and Technology), it was found that there were no significant differences between subgroups reading an authentic (in original, unadapted form) computer science text and those reading a syntactically adapted text either in comprehension or in time across the four groups of subjects. These results indicate that syntactic simplification of an EST text is not a real simplification. Hence, technical writers and EST teachers might give priority to other more conceptual ways of rewriting texts.
simplification syntaxique d'un texte scientifique/technique; 2) la connaissance de l'arrière-plan par le lecteur (expert/novice) sur l'analyse syntaxique d'un texte; et 3) la connaissance qu'a le lecteur du langage (L1, anglais langue maternelle/L2, anglais langue seconde) sur l'analyse syntaxique d'un texte. Il se peut que, pour un lecteur expert ou novice dans un champ scientifique ou technique particulier, ici l'informatique, la compréhension du texte exige d'effectuer une analyse conceptuelle et lexicale complète mais pas une analyse syntaxique exhaustive.

Les sujets de l'expérience étaient des étudiants américains (L1) et hollandais (L2) spécialisés en Informatique (experts) et en Lettres (novices) ayant à lire un article de journal scientifique relatif à l'informatique. On a choisi, pour les réécrire, dix passages de une phrase, également répartis dans tout le texte. Ces phrases, qui contenaient des nominalisations, des passifs, et des constructions de participes pertinentes pour le registre de l'EST (English for Science and Technology), ont été réécrites dans une forme syntaxique plus adaptée à un registre de langage commun. On a pris soin de ne pas changer le sens des phrases. Aucun item lexical ou autre item des phrases d'origine n'a été changé; par conséquent on a maintenu la même charge de contenu.

Les résultats de la recherche ont montré que le registre de langue du texte n'a eu aucun effet significatif sur les scores de compréhension de lecture, mais que la connaissance du contexte et la connaissance du langage ont eu un effet (p <0.01; p <0.005). Partant de là, on peut conclure qu'il n'y a pas de différence significative dans la compréhension d'un texte authentique par rapport à celle d'un texte réécrit (simplifié), mais que la connaissance de l'informatique aide les sujets à comprendre significativement mieux.

Dans cette recherche, les sujets ont enregistré eux-mêmes leurs temps de lecture et quoique ces temps ne soient peut-être pas statistiquement fiables, ils apportent toutefois une information superficielle. Il y a un effet très important de la connaissance de la langue sur les temps de lecture (p <0.005). Ce facteur intervient dans trois interactions hautement significatives: 1) avec la connaissance du contexte (p <0.005), avec le registre de langue (p <0.005), et 3) avec la connaissance du contexte et le registre de langue (p <0.005). Ces interactions sont dues au fait que les Américains (L1) lisent deux fois plus vite que les Hollandais (L2). La connaissance du contexte informatique (experts/novices) augmente bien la compréhension, mais elle ne diminue pas le temps de lecture.

L'adaptation syntaxique du texte informatique anglais n'a pas réellement aidé les étudiants à mieux comprendre ou à lire plus vite, qu'il s'agisse des Américains ou des Hollandais, même s'ils manquaient de connaissances en informatique.

Cette étude a d'importantes implications. La façon dont les gens lisent un texte dans leur langue maternelle ou dans une langue étrangère a un impact sur la façon dont le texte est écrit. Plusieurs expériences ont confirmé qu'il n'est pas nécessaire de réécrire la syntaxe de textes professionnels pour augmenter leur lisibilité. Ils semblerait plus avantageux, par contre, d'effectuer une révision textuelle ou lexicale. Pour les professeurs de EST, il est nécessaire d'aider les étudiants à développer des stratégies de lecture de textes à de hauts niveaux de compréhension en mettant l'accent sur le développement d'habiletés textuelles pour acquérir l'information requise des textes, et sur le développement d'un vocabulaire technique et sous-technique plutôt que de se focaliser sur une analyse syntaxique précise.
INTRODUCTION

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, 1949; Fry, 1968) account for surface elements in a text such as sentence length and word length. Since readability formulas are designed to analyze prose text, problems arise when these formulas are applied to scientific, technical, and mathematical material, for these contain numerical and symbolic language and specialist vocabulary (Hartley, 1985). Some syntactic variables are referred to in many guidelines which are designed to help writers produce readable material. Writers are told 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. There is the suggestion that the use of nominalizations (converting verbs into nouns) makes sentences more complex because they are "indirect" (e.g., Klare, 1985; Price, 1984: 143). There are, of course, many other elements that affect readability. Some of these are vocabulary, cohesion, paragraph length and organization, concreteness or imageability, referents, and reader motivation. There now seems to be consistent research evidence against using readability formulas to measure success in writing and rewriting technical and legal documents for adult readers (e.g., Duffy, 1985).

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 (Binkley, 1988). These systems tend to focus on text factors that influence learning and memory rather than text factors which correlate with reading difficulty. While they have contributed to general understanding of the reading process in general, they do not specifically address the contribution of individual syntactic features to readability of texts.

Studies on the characteristics of English for Science and Technology (e.g. Trimble, 1985; Ulijn, 1979, 1984) 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). In a recent experiment (Strother & Ulijn, in press) both American and Dutch computer science and humanities students significantly preferred to use this ST syntax in a computer science (CS) text and CL syntax in a fiction text. It might be assumed then that knowledge of and/or increased practice with specific ST structures would improve a student's reading comprehension level of ST texts.

Indeed, some of these specific syntactic structures are 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. Moreover, Charrow and Charrow (1979) found that some structures, such as subordinate passives, as to phrases, and nominalizations, caused problems in American jury instructions. Bhatia (1984) ascertained similar 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 [authors’ note: 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.

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.

Kieras (1985) used referential/anaphoric, syntactic and textual comprehensibility rules to revise a bad version of a simulated technical manual and found significant improvement in reading and operating time and success, but what was the specific contribution of the syntactic factor to this? Experiments for L1 indicate that syntactic “simplification” is not always a real simplification. Is this the same for EST in an L2 setting? The need for simple clear language in both L1 and L2 situations is obvious.

In 1978, US 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). In Britain, there is also a Plain English Campaign, with its headquarters in Stockport. That organization publishes newssheets, 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 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. This answer is also relevant to EST teachers who wish to adapt original specialist and lay materials for second language learners. In fact, we are following the call for experimental research by Davis and Kantor (1982) to determine if syntax really affects readability as opposed to the implicit suggestion of some readability formulas that syntax does so.

In contrast, present cognitive psychology and psycholinguistics emphasize schema-based, conceptually guided, interactive, and partially parallel strategies to process texts. (For a good description of the current schema theory, see Rumelhart,
This implies that high 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. The limited importance of syntax for reading is strongly evidenced by Ulijn (1981), but those experiments dealt with French instructions written in a common rather than a scientific register. The experiments reported on here were designed to replicate the above study specifically for EST. If the same results were found in this replication, it would demonstrate that syntactic adaptation of an EST text would not really help a reader comprehend a text better. Hence, technical writers and EST teachers might give priority to other more conceptual ways of simplifying texts.

Certainly, a knowledge of syntax is a requirement for reading comprehension. According to most current experimental literature, however, syntax plays only a minor role in both native and second language reading comprehension at an advanced level. 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 E2 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).

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). 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.

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 & Swales (1984); Alderson & Urquhart (1985); Peretz & Shoham (in press). 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 cognition processing changes too (Kintsch & Miller, 1984; Zakaluk & Samuels, 1988).

So far, the possible effects of syntax, background and language knowledge (E1 and E2) on readability and reading comprehension have been discussed, which leads to the specific questions of this experiment.
Research Questions

The research questions to be addressed include the following:

1. Is syntactic simplification a real simplification of English texts, in particular those written for the domain of science and technology (EST)?

The general result, however, could be contaminated by topic familiarity and language knowledge. Hence, we add two questions:

2. To what extent does background knowledge affect syntactic analysis in reading an EST text?
3. To what extent does language knowledge affect syntactic analysis in reading an EST text?

Therefore, there are three factors in the study: (1) the register of the text — either ST (science and technology — unsimplified) or CL (common language — simplified); (2) the background knowledge of the readers — either CS (computer science) experts or nonexperts; and (3) the background language of the readers — either L1 or L2.

METHOD

Subjects

Four groups of subjects were involved in the study:

(1) 24 native computer science majors in the United States.
(2) 24 native humanities majors in the United States.
(3) 24 nonnative (Dutch) computer science majors in The Netherlands.
(4) 24 nonnative (Dutch) humanities majors in The Netherlands.

Table 1 shows a summary of subjects by native language, background knowledge of computer science (CS) and language register (ST or CL) of the text to be read, a $2 \times 2 \times 2$ experimental design.

These subjects were from three universities: Florida Institute of Technology in Melbourne, Florida (US), Eindhoven University of Technology and Tilburg University in The Netherlands. Those subjects — both L1 and L2 — who were judged to be, for the purposes of this study, experts in CS were in their third or fourth year of study in a CS major at university. The nonexperts in CS were majoring in humanities (such as language and literature, law, and economics) at university and ascertained that they did not have formal training or background knowledge in the CS field. Fourteen of the twenty-four humanities majors were in their first year.

The subjects in the American universities all had English as a native language. The Dutch subjects were all native speakers of Dutch and had a strong English background: six years of English at the secondary school level, sometimes with additional university course, and with considerable exposure to English textbooks during their three years of study (a high intermediate to advanced level). Their English knowledge
Table 1. Number and distribution of subjects over the experimental conditions

<table>
<thead>
<tr>
<th>Language Knowledge</th>
<th>American (E1)</th>
<th>Dutch (E2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Register:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authentic (A)(ST)</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Syntactically adapted (B)(CL)</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Background Knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Science Majors (experts)</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Humanities Majors (novices)</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>TOTALS</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

Registers:
ST = Science and Technology
CL = Common Language

was not near native (TOEFL range 475–550 as a point of reference; see Perkins, 1988), and certainly was not similar to that of the American subjects. Hence, their syntactic knowledge should be different as well. The Dutch also had a traditional weaker schooling in French and German, whereas the Americans had no specific foreign language knowledge. We can safely conclude that the effects of language knowledge other than English and Dutch will not contaminate our design and that the English syntactic knowledge of Dutch and Americans is not alike.

Materials and Procedure

A computer science article, which had no mathematical formulas or illustrations, was chosen for the text (Miller, 1982). Ten one-sentence passages, equally distributed over the entire text, were chosen for rewriting. They contained nominalizations, passives, and participle constructions pertinent to the EST register and mainly meant to make the agent secondary to the action (See Appendix B for a full list.) These sentences were rewritten in a syntactic form more suitable for a common language (CL) version. For example:

Authentic version: Replacement at a higher system level is required; this implies considerable software development to take advantage of these promising device characteristics.

Syntactically revised version: 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.

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 copy of the article and the question set were checked by an expert in the computer science field to ascertain 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. Two types of test booklets were made: Test Booklet A used the original (authentic) article and Test Booklet B used the partly rewritten version. Both were followed by the same 10 true-false statements (formulated as much as possible in a text-independent way) referring in random order to the 10 passages that had been rewritten in the text. A list of the original and rewritten passages from the CS text is given in Appendix A. Appendix B shows to what extent the two versions syntactically differ (sentence length and types of linguistic structure) and how the answer sheet relates to the 10 test sentences (the statement order of the answer sheet is scrambled). Syntactic "simplification" appeared to require on the average slightly more words than the original versions (303 vs. 296), but four simplified sentences (numbers 3, 4, 5, and 8) were shorter. In general, we tried to eliminate a possible effect of different sentence length as a contaminating factor.

Subjects were randomly assigned to one of the two test forms and were asked to record their beginning and ending times on the test booklets.

Comment on the Procedure Used

We realize that as a test, our procedure could be weak for two reasons:

1. True/False items encourage guessing, but another statistically more reliable answering technique would be difficult to develop in a setting which should be as real-life as possible.
2. The framework of an original CS text did not allow us to vary the syntactic features of the ST register in a very systematic way and to make a valid check on item difficulty. To avoid artificial item descriptions, the sentences were used exactly as they appeared in the original computer science 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). We consider the adopted procedure to be adequate for our research question.

Design and Statistics

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

1. Language register of the text: (LR)
   1. Original version (ST)(A)
   2. Syntactically adapted version (CL)(B)

2. Background knowledge (BK)
   1. Computer science (experts)
   2. Humanities (novices)

3. Language knowledge (LK)
   1. Native English (American)
   2. Nonnative (Dutch)

In addition, we calculated some two-way and three-way interactions. Additional data came from subjects with other language backgrounds (Strother & Ulijn, 1987). The dependent variables were (1) number of answers correct (out of a maximum score of ten) on the comprehension test and (2) text reading time.
RESULTS

As can be seen in Table 2 and Figure 1, language register did not have a significant effect on reading comprehension scores, but both background knowledge and language knowledge did (p < .01 vs. p < .005). From that we can conclude that there are no significant differences in comprehension between versions A and B, but that knowledge of computer science did help the subjects to comprehend significantly better. Moreover, Dutch students had very significantly higher comprehension scores than the Americans.

Table 2. ANOVA results

<table>
<thead>
<tr>
<th>Main Effects of:</th>
<th>on Comprehension Scores</th>
<th>on Reading Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(1.33)</td>
<td>p</td>
</tr>
<tr>
<td>LR: Language register of the text:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original vs. Syntactically Adapted</td>
<td>&lt;1</td>
<td>n.s.</td>
</tr>
<tr>
<td>BK: Background knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Science vs. Humanities</td>
<td>9.94</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>LK: Language knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native vs. Nonnative English</td>
<td>21.12</td>
<td>&lt;.005**</td>
</tr>
</tbody>
</table>

Interactions

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LK x BK</td>
<td>&lt;1</td>
<td>n.s.</td>
<td>&lt;1</td>
<td>n.s.</td>
</tr>
<tr>
<td>LK x LR</td>
<td>&lt;1</td>
<td>n.s.</td>
<td>20.43</td>
<td>&lt;.005**</td>
</tr>
<tr>
<td>LK x BK x LR</td>
<td>&lt;1</td>
<td>n.s.</td>
<td>22.24</td>
<td>&lt;.005**</td>
</tr>
</tbody>
</table>

* significant
** highly significant

According to Figure 2, there is a very important effect of language knowledge on reading times (p < .005). This factor is involved in three highly significant interactions: (1) with background knowledge (p < .005), (2) with language register (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. Since reading times were recorded by the students themselves and had a very high standard deviation in each cell of observations, reading time was a very unreliable factor. Therefore, we are not tempted to draw substantial conclusions from these interactions. Reading time was just meant to be a superficial check on comprehension. For this reason, no correlation coefficient was calculated between comprehension and time.

DISCUSSION AND CONCLUSION

Ulijn (1981) reported research suggesting a conceptual strategy of foreign language reading: nonnative readers were hampered by all kinds of lexical problems since they
Figure 1. Mean Comprehension Scores (Numbers of Correct Answers)

<table>
<thead>
<tr>
<th>Number of Correct Answers</th>
<th>E1</th>
<th>E2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.58</td>
<td>9.67</td>
</tr>
<tr>
<td>9</td>
<td>8.25</td>
<td>9.33</td>
</tr>
<tr>
<td>8</td>
<td>8.17</td>
<td>8.75</td>
</tr>
<tr>
<td>7</td>
<td>7.25</td>
<td></td>
</tr>
</tbody>
</table>

Key:
- **LANGUAGE REGISTER:**
  - • Authentic version (ST)
  - ○ Syntactically adapted version (CL)
- **BACKGROUND KNOWLEDGE:**
  - ——— Computer Science Majors
  - ——— Humanities Majors

Figure 2. Mean Reading Times (in minutes)

<table>
<thead>
<tr>
<th>Reading time in minutes</th>
<th>E1</th>
<th>E2</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>17.33</td>
<td>16.42</td>
</tr>
<tr>
<td>16</td>
<td>14.33</td>
<td>13.58</td>
</tr>
<tr>
<td>14</td>
<td>10.83</td>
<td>9.50</td>
</tr>
<tr>
<td>12</td>
<td>9.25</td>
<td></td>
</tr>
</tbody>
</table>

Key:
- **LANGUAGE REGISTER:**
  - • Authentic version (ST)
  - ○ Syntactically adapted version (CL)
- **BACKGROUND KNOWLEDGE:**
  - ——— Computer Science Majors
  - ——— Humanities Majors
could not avoid analyzing the conceptualization of the text based upon its vocabulary. Syntactic aspects were not difficult because they seemed to remain unnoticed by the nonnative reader. In this case, the experiment was done with French and Dutch subjects reading French instructions to find their way through an imaginary town. The new study was conducted to determine what would happen to English read by natives (Americans) and nonnatives (Dutch and a number of other language groups).

Indeed the results gathered by Strother and Ulijn (1987) for Chinese, Arabic, Spanish, and others (with varying numbers) were largely confirmed in the above $2 \times 2 \times 2$ setting with equal numbers in the 8 cells of concern (see Table 1). Both native and nonnative readers of an English computer science text seemed to overlook all kinds of syntactic variants that have usually been assumed to be simplifications: nominalizations vs. verb phrases, passive vs. active constructions, and participle constructions vs. subordinate clauses. In the current experiment, the adaptation from a scientific and technical register syntax into a more simplified common language syntax did not affect either the comprehension scores or the reading time.

Background knowledge of computer science (experts vs. novices) did increase comprehension, but it did not decrease reading time. This seems to correspond with recent research on E2 engineering and social science texts (Shoham, Peretz, & Vorhaus, 1987). Nonnative knowledge of English (by the Dutch) allowed for a better comprehension than native knowledge (by the Americans), but in a longer reading time. As could be expected, native readers took less time than nonnative readers. For native reading, it has been experimentally shown that higher comprehension scores do correlate with shorter reading times (Jackson & McClelland, 1975) and that there is an optimal reading rate which is constant throughout high school and college ages and which does not necessarily correlate with text difficulty (Carver, 1983). Our data support Berkoff's suggestion (1979) that a fast reader is not always an efficient reader and a slow reader not always an inaccurate one, in particular for nonnative reading. The English text might tend to have a signalling effect on Dutch readers; it might catch their attention more than the same text does for American subjects and therefore leads to a more accurate but more time-consuming comprehension. The Dutch were superior in comprehension but were not in reading time.

The syntactic adaptation of the English computer science text did not really help the students to comprehend better or to read more quickly, either for the Americans (natives) or for the Dutch (nonnatives), even if they lacked knowledge of computer science. However, there are some tendencies, which, while not statistically significant, are worth mentioning. Comprehension for both language groups is slightly better with syntactic adaptations of the text for both experts and novices, for both native and nonnative readers, with the exception of Dutch humanities majors. They comprehended the original text as well as the simplified text. Moreover, one should warn against overgeneralization of this apparent lack of syntactic effect. The experimental test might not be sensitive enough to demonstrate it because the comprehension scores for both native and nonnative readers are very high; the Dutch CS majors scored 9.33 and 9.67 out of a possible 10.0 on the original and adapted versions. Therefore, we can assume a ceiling effect in our data. There is not much scope for an increase in the scores of the nonnative CS majors. The test does not discriminate enough between native and nonnative reading scores. Moreover, comprehension score was a more reliable factor than reading time.
It also 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. The ceiling effect might mask a possible syntactic effect on the Dutch readers because the Dutch and American groups were too alike in their science-based English knowledge. But this is contradicted by the fact that general language knowledge (native vs. nonnative English) affected the reading comprehension scores and times at a highly significant level, as would happen to lower proficiency language users than the Dutch. Even with these tendencies, these results [confirmed by those of other language groups (Strother & Ulijn, 1987)] support the idea that neither native nor nonnative readers of EST texts need a thorough syntactic analysis. Therefore, at advanced levels, syntactic simplification into a more common language register does not really increase readability. Background knowledge, however, is a significant independent factor in both native and nonnative reading comprehension which gives way to a rather lexicon-based conceptual strategy of reading (not tested here).

Implications

What, then, are the implications for these findings? First, for textbook authors and technical writers, the implications are noteworthy. The way in which people read a text in their native language or a foreign language should have an impact on how the text should be written. In a number of experiments, the conclusion is that syntactic rewriting of professional texts to increase readability is not needed. The sentence is not a good unit for rewriting. As has become apparent from earlier experiments and analyses (Bouwman et al., 1985; Bovair & Kieras, 1985) textual (propositional) and lexical rewriting might have much more effect.

Readability measures should include careful manipulation of cohesive structures of a text, as suggested by Olsen and Johnson (in press) in their critique of Duffy and Kabance’s results. 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 Beck et al. (1984) for second grade material. This is not to deny a syntactic threshold which should be overcome when one starts learning to read in a foreign language as Alderson (1984) suggests and Stone’s (1985) experimental results confirm. In advanced reading comprehension, however, simplification seems to be a matter of textual and lexical revision rather than of syntactic revision.

The results are important for EST teachers to apply. There is the need to help students develop strategies for reading texts with high levels of comprehension. Instead of focusing on syntax, the focus should be on development of 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 E2 reading comprehension, as has been shown by Bensoussan (1987). Students 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 certain syntactic structures occur with greater frequency in EST texts, students can be helped to focus on these structures, especially those that seem to give students more trouble. Also, they can be helped to fill in any gaps they have in these particular structures. Hopefully, then, writers can improve the readability of
materials students are given to read and instructors will be better able to help students improve their reading comprehension of academic materials.

Since lexical and textual simplifications were not tested in the present experiment, these aspects call for detailed verification before drawing any final conclusion about their simplifying effect. The experimental results reported in the Introduction on the significant role of background knowledge in both native and nonnative reading, however, is confirmed by one of our present findings. This aspect should, therefore, be effectively used in the teaching of EST reading.

ACKNOWLEDGEMENTS

This experiment was a cooperative effort between Florida Institute of Technology and Eindhoven University of Technology. The authors are indebted to Dr. Thomas Hand, Professor of Computer Science at Florida Institute of Technology for his help with checking the syntactic rewriting of the text, verifying the questions, and confirming the equivalence of both versions, and to Mr. Jelle Buizer and Mr. Lennard Peeters, Technology and Communication students at Eindhoven, for their experimental and statistical assistance.

REFERENCES


SYNTACTIC SIMPLIFICATION AND EST READING


Address for Correspondence: JAN M. ULIJN, Section of Psychology and Language, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

APPENDIX A

Sentence | Authentic version (A) [in EST (English for Science and Technology) register]
--- | ---
1. **Replacement** at a higher system level is required; this implies considerable software **development** to take advantage of these promising device characteristics.
2. Certainly, the 1/0 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.
3. 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.
4. 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.
5. 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.
6. During the last three 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.
7. 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.

8. The data center manager who is concerned about reducing operating costs or improving performance of this central repository may contemplate introducing optical data disks.

9. When such movement is possible under system control, the problem becomes one of separating and clustering sets of files independently by the physical volumes and providing a mechanism to have desired files more readily available.

10. In handling more tape faster, many data center personnel must mount manually hundreds to thousands of tape reels per day.

Syntactically revised version (B)

1. 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.

2. Certainly the 1/0 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.

3. These constitute “clean interfaces” because the user may implement something now to replace a single or small set of layers (modules) without having to redesign (or emulate) the entire set.

4. 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.

5. The term “backfill staging” indicates that the host can direct such a transfer without moving data into and out of the main memory.

6. During the last three 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.

7. 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.

8. A data center manager may introduce optical data disks if he must reduce operating costs or improve performance of this central repository.

9. 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.

10. Because the system must handle more tape faster, many data personnel must mount manually hundreds to thousands of tape reels per day.
APPENDIX B

Syntactic Analysis of Authentic Text (A) Versus Syntactically Adapted Text (B)

<table>
<thead>
<tr>
<th>Sentence Number</th>
<th>Number of Words</th>
<th>Linguistic Structure</th>
<th>Statement number on Answer Sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>29</td>
<td>2 × nominalization</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>31</td>
<td>nominalization</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>30</td>
<td>nominalization</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>29</td>
<td>passive</td>
</tr>
<tr>
<td></td>
<td>1 sentence</td>
<td>2 sentences</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>23</td>
<td>passive</td>
</tr>
<tr>
<td></td>
<td>nominalization</td>
<td>verb</td>
<td></td>
</tr>
<tr>
<td></td>
<td>participle</td>
<td>clause</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>32</td>
<td>participle</td>
</tr>
<tr>
<td></td>
<td>nominalized verb</td>
<td>clause</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>45</td>
<td>48</td>
<td>participle</td>
</tr>
<tr>
<td></td>
<td>nominalized verb</td>
<td>clause</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>22</td>
<td>3 × participle</td>
</tr>
<tr>
<td>9</td>
<td>35</td>
<td>36</td>
<td>3 × participle</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>23</td>
<td>participle construct</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td>276</td>
<td>303</td>
<td></td>
</tr>
</tbody>
</table>