J. M. Ulijn, J. B. Strother

Is Syntactic Simplification a Real Simplification of EST Texts?

Series C: Language of Specific Purposes
ISSN 1435-6473
Essen: LAUD 1987 (2nd ed. with divergent page numbering 2013)
Paper No. 14

Universität Duisburg-Essen
J. M. Ulijn, J. B. Strother

University of Eindhoven (The Netherlands), Florida Institute of Technology (US)

Is Syntactic Simplification a Real Simplification of EST Texts?

Copyright by the author
1987 (2nd ed. with divergent page numbering 2013) Reproduced by LAUD
Series C Linguistic Agency
Language of Specific Purposes University of Duisburg-Essen
Paper No. 14 FB Geisteswissenschaften

Order LAUD-papers online: http://www.linse.uni-due.de/linse/laud/index.html
Or contact: laud@uni-due.de
Is Syntactic Simplification a Real Simplification of EST Texts?

There has been a considerable suggestion in the literature that syntactic adaption of a text would simplify it, for instance, a science & technology text written in English.

At Florida Institute of Technology in the US and at Eindhoven University of Technology in The Netherlands, an empirical analysis was done to determine that the complexity of the syntax does not significantly affect the level of reading comprehension. While a complete conceptual and lexical analysis is necessary for reading comprehension, a thorough syntactic analysis is not. This should true for both expert and novice readers in a particular professional field, for example that of computer science.

There were four groups of subjects:
1. native computer science majors in the US
2. native humanities majors in the US
3. non native (Dutch) computer science majors in The Netherlands
4. non native (Dutch) humanities majors in The Netherlands.

It was found that there were no significant differences between subgroups reading an authentic computer science text and those reading a syntactically adapted text neither in comprehension nor in time across the four groups of subjects. However, non native readers took significantly more time to comprehend better than native readers. It is also true that background knowledge of computer science did increase significantly comprehension.

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.

1. Introduction

For the past 30 or 40 years much attention has been paid to formulas that claim to measure readability. Most of these formulas, such as the well known Flesch one (1949) account for surface elements in a text such as sentence length and word length. Many guidelines to produce readable writing refer to syntactic variables, such as “keep your sentences active” (avoid passives and nominalizations) and “make separate sentences of dependent clauses” (Harrison 1980: 23). The question is: “Is syntactic simplification a real simplification of English texts, in particular those written for the domain of science and technology (EST)?”

Because of the “Plain English” law 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 layman documentation, for example, in computer science and
information technology. This answer is also relevant to EST teachers who wish to adapt authentic specialist and layman materials for second language learners. In fact, we are following the call for experimental research by Davison & Kantor (1982) to check if syntax really affects readability, as opposed to the implicit suggestion of some readability formulas that syntax does so.

Present cognitive psychology and psycholinguistics suggest a preference for schema-based, conceptually guided and interactive partially parallel strategies to process texts (for a good description of the current schema theory, see Rumelhart 1981). 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 unimportance of syntax for reading is strongly evidenced by Uljın (1981), but his experiments deal with French instructions which are rather written in a common register. The experiments we report on here are designed to replicate the above idea for EST. If this is true, syntactic adaption of an EST text would not really help. Hence, technical writers and EST teachers might give priority to other more conceptual ways of simplifying texts.

Authentic texts are often thought to be more difficult to comprehend than simplified ones. One reason for this is the complexity of the syntax in authentic texts. Certainly, a knowledge of syntax is requirement for reading comprehension. Nilagupta (1977) and Alderman & Alvarez (1978) present some evidence on that score. According to most current experimental literature, however, syntax plays only a minor roll in both native and second language reading comprehension. Garrod (1984), just to mention more recent studies, reports that the lowest factor contributing to variation in reading time in E, was syntactic complexity. In reading EST (E₂) texts Venezuelan science students experience less syntactic than lexical difficulties (Akirov & Salager 1984). This is in line with the results of Isreali science & technology students whose syntactic knowledge seems to have contributed the least to their E₂ reading comprehension (Weiss 1985).

There have been an increasing number of studies on the characteristics of English for Science and Technology (see, for example, Todd Trimble et al. 1978; Uljın 1984a and 1985). These studies show that there are some syntactic structures that appear with greater frequency in scientific and technical text than they do in text which is written in the common register. It might be assumed then that knowledge of these specific structures would improve a student’s reading comprehension level.

Some of these specific syntactic structures are known to cause difficulty in comprehension. For example, Charrow and Charrow (1979) found that some structures, such as subordinate passives, as to phrases, and nominalization, caused problems in American jury instructions. Bathia (1984) ascertained similar syntactic problems in British legislative writing. The English judicial register might have a different syntactic impact on reading than the EST register would have.
2. Hypotheses

This research takes structures that occur with greater frequency in the EST register (Ulijn 1984a) and simplifies those structures. The structures that are rewritten here are passives, nominalizations, and participles. The experiment was conducted to test the validity of the hypothesis that no significant differences will be obtained between scores on a reading comprehension test by university students who read an unsimplified and those who read a syntactically simplified computer science journal article. However, one could expect that syntactic simplification would be a real simplification for students without background knowledge of the text’s subject matter or for students without a native knowledge of English, since these readers would have to revert to a more detailed syntactic analysis.

Therefore, differences in scores among the following groups were examined: computer science majors and humanities majors whose first language is English (E₁), and computer science and humanities majors first language is Dutch and whose second language is English (E₂).

3. Method

3.1. Subjects

The subjects of the study were from three universities, Florida Institute of Technology in Melbourne, Florida (US), Eindhoven University of Technology and Tilburg University in The Netherlands, and included both undergraduate students in their third or fourth year and beginning graduate students. Fourteen of the 24 humanities majors were in their first year. Table 1 shows a summary of our subjects by native language, background knowledge of computer science (CS), and language register of the text to be read (a 2 x 2 x 2 experimental design).

<table>
<thead>
<tr>
<th>3. Language Knowledge</th>
<th>3.1. American (E₁)</th>
<th>3.2. Dutch (E₂)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Language Register:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1. Authentic (A)</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1.2. Syntactically (B) adapted</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2. Background Knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1. Computer Science Majors (experts)</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>2.2. Humanities Majors (novices)</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Totals</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1: Number and distribution of subjects over the experimental conditions

The 48 American students and the 48 Dutch students were half CS majors in each group and the other half were humanities majors with no formal training in CS.
The Dutch students had a comparable English background: six years of English at the secondary school level, sometimes an additional university course, and a considerable exposure to English textbooks during their three years of study. (A high intermediate to advanced level).

3.2. Materials

A computer science article, which had no mathematical formulas or illustrations, was chosen for the test (Miller 1982). Ten one-sentence passages, equally distributed over the entire text, were chosen for rewriting. They contain nominalizations, passives and participle constructions pertinent to the EST register and mainly meant to make the agent secondary to the action. These sentences were rewritten in a syntactic form more suitable for a common language (CL) version (see fig. 1). 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 authentical article and Test Booklet B used 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.

3.3. Procedure

Subjects were randomly assigned to one of the two test forms and were asked to record the time spend on reading the text.

2.4. Design & Statistics

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

1. Language Register of the text:
   1. Authentic version (EST) (A)
   2. Syntactically adapted version (B)

2. Background Knowledge:
   1. Computer Science (experts)
   2. Humanities (novices in CS)

3. Language Knowledge:
   1. Native English (American)
   2. ESL/EFL (Dutch)

Additional data came from subjects with other language backgrounds (Strother & Ulijn, in press). The dependent variables were number of answers correct on the comprehension test and text reading time.
4. Results

Figure 2 shows the comprehension scores.

Language register did not have a significant effect on these scores, but both background knowledge and language knowledge did, $F(1,88) = 9.94, p < .01$ significant and $F(1,88) = 21.12, p < .005$, highly significant.

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 to comprehend better. Moreover, Dutch students understood the text better than the Americans.

According to Fig. 3 there is a main effect of language knowledge on reading times, $F(1,88) = 18.07, p < .005$, highly significant.
This factor is involved in 3 highly significant interactions:

- X background knowledge, $F(1,88) = 20.43$, $p < .005$
- X language register, $F(1,88) = 18.18$, $p < .005$
- X b. k. X l. r., $F(1,88) = 22.24$, $p < .005$

There are due to the fact that Americans read nearly twice as quickly as Dutch. Since reading time was a very unreliable factor (recorded by the students themselves and with a very high standard deviation in each cell of observations), we are not tempted to draw substantial conclusions from these interactions.

5. Discussion

The conceptual strategy of reading proposed by Ulijn (1981) and demonstrated by French and Dutch subjects reading French instructions to find their way through an imaginary town (Beausite) has been largely confirmed: both native and non-native readers of an English computer science text mainly adopt a conceptual strategy aiming at content words and therefore overlook all kinds of syntactic variants that have usually assumed to be simplifications: nominalizations vs. verb phrases, passive vs. active constructions and participle constructions vs. subordinate clauses. In our experiment, the adaptation from a
scientific and technical register syntax into a more simplified common language syntax neither affected the comprehension nor the reading time.

Background knowledge of computer science (experts vs. novices) did increase comprehension, but it did not decrease reading time. Non-native knowledge of English (of the Dutch) allowed for a better comprehension than native knowledge (of the Americans), but in a longer reading time. As could be expected, native readers took less time than non-native readers. This does correspond with the current finding that in efficient reading higher comprehension scores do correlate with shorter reading times.

The syntactic adaption of the English computer science text did not really help the students to comprehend better or to read more quickly, neither for the Americans (natives) nor for the Dutch (non-natives), even if they missed knowledge of computer science.

However, there are some tendencies, which, while non-statistically significant, are striking. Comprehension for both language groups is slightly better with syntactic adaptations of the text for both experts and novices, for both native and non-native 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. This test might not be sensitive enough to demonstrate it, because the comprehension scores for both native and non-native readers are very high: there is a ceiling effect. Even with these tendencies, these results (confirmed by those of other language groups, see Strother & Ulijn, in press) support a conceptual strategy used by both native and non-native readers of EST texts: a thorough syntactic analysis is not needed. Therefore, syntactic simplification into a more common language register does not really increase readability.

6. **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 what is needed is not syntactic rewriting of professional texts to increase readability. The sentence is not a good unit for rewriting. As has become apparent from recent experiments (Kieras, 1982, Bouwman et al. 1985 and Bouwman & Ulijn, in preparation) textual and lexical rewriting will have much more effect. For researchers, the hypothesis of partial parallelism between conceptual and syntactic analysis in reading (Ulijn 1981 and 1984) has been confirmed for English as a native and a second language.

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 development of both the technical and subtechnical
vocabulary. We should guide our students toward focusing more on concepts and on the vocabulary, with syntactic analysis being superficial in most cases. Hopefully, then, we can improve the readability of materials we give our students to read and will be better able to help our students improve the reading comprehension of their academic materials.

7. Acknowledgements

This experiment is the fruit of American-Dutch cooperation between Florida Institute of Technology and Eindhoven University of Technology. The authors are indebted to Dr. Tomas Hand 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 for their experimental and statistical assistance.

8. Authors

Jan M. Ulijn, a researcher in psycholinguistics, second language acquisition, and language for specific purposes, teaches technical communication theory at Eindhoven University of Technology in The Netherlands.

Judith B. Strother, specializing in reading English for Science and Technology and English for Academic Purposes, teaches at Florida Institute of Technology in Melbourne, Florida (US).
References


<table>
<thead>
<tr>
<th>Sentence</th>
<th>Authentic version (A) (in EST* register)</th>
<th>Syntactically revised version (B) (in CL ** register)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><em>Replacement at a higher system level</em> is required; this implies considerable software <em>development</em> to take advantage of these promising device characteristics.</td>
<td>The user must <em>replace</em> current devices with optical data disks at a higher system level. The programmer must <em>develop</em> considerably more software to take advantage of optical data disks.</td>
</tr>
<tr>
<td>2.</td>
<td>Certainly the I/O structure of the present day computer does not constitute a clean interface, primarily because the operating systems assume <em>knowledge</em> of the physical characteristic of the device.</td>
<td>Certainly the I/O structure of the present day computer does not constitute a clean interface, interface, primarily because the operating systems assume that the use knows the physical characteristics of the device.</td>
</tr>
<tr>
<td>3.</td>
<td>These constitute “clean interfaces” because a brand-new <em>implementation</em> may be introduced to replace a single or small set of layers (modules) without having to redesign (or emulate) the entire set.</td>
<td>These constitute “clean interfaces” because the user may <em>implement</em> something now to replace a single or small set of layers (modules) without having to redesign (or emulate) the entire set.</td>
</tr>
<tr>
<td>4.</td>
<td>Via telephone lines, leased lines, concentrators etc., remote terminals are <em>connected</em> to a front-end processor, which often permits access to one of several host-processors at the data center.</td>
<td>Telephone lines, leased lines, concentrators, etc., <em>connect</em> remote terminals to a front-end processor. This processor often permits access to one of several host processors at the data center.</td>
</tr>
<tr>
<td>5.</td>
<td>The term “backfill staging” has been used to indicate the possibility of the host <em>directing</em> such a transfer without the movement of data into and out of the main memory.</td>
<td>The term “backfill staging” indicates that the host can direct such a transfer without moving data into and out of the main memory.</td>
</tr>
<tr>
<td>6.</td>
<td>During the last three decades, <em>considering the density of storage</em> on the tape and the speed with which we move tape, tape technology has become only a few</td>
<td>During the last three decades, <em>if we consider that the tape stores more data</em> in a smaller space and moves much faster, tape technology has become only a few hundred times</td>
</tr>
</tbody>
</table>
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.

The meaning of “getting the architecture straight” may be debatable; however, the act of getting the architect 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.

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

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.

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

---

* EST: English for Science & Technology
** CL: Common Language

Fig. 1: List of 10 sentences from the computer science text used in the experiment, pertinent to the EST register and syntactically revised in a rather CL version, with number of words, distributed over types of linguistic structures and answer sheet statements.
<table>
<thead>
<tr>
<th>Number of words</th>
<th>Linguistic structure</th>
<th>Statement number of answer sheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 21, B 29</td>
<td>2x nominalization</td>
<td>5.</td>
</tr>
<tr>
<td>A 29, B 31</td>
<td>nom. verb</td>
<td>1.</td>
</tr>
<tr>
<td>A 31, B 30</td>
<td>nom. verb</td>
<td>8.</td>
</tr>
<tr>
<td>A 30, B 29</td>
<td>passive constr. nom. verb</td>
<td>2.</td>
</tr>
<tr>
<td>A 30, B 23</td>
<td>passive constr. nom. participle constr</td>
<td>3.</td>
</tr>
<tr>
<td>A 31, B 32</td>
<td>part. constr. nom. verb</td>
<td>9.</td>
</tr>
<tr>
<td>A 45, B 48</td>
<td>part. constr. nom. verb</td>
<td>7.</td>
</tr>
<tr>
<td>A 24, B 22</td>
<td>3x part. constr. clauses</td>
<td>4.</td>
</tr>
<tr>
<td>A 35, B 36</td>
<td>3x part. constr. clause</td>
<td>10.</td>
</tr>
<tr>
<td>A 20, B 23</td>
<td>part. constr. clause</td>
<td>6.</td>
</tr>
</tbody>
</table>