Interlanguage and EST Writing: Some Syntactic Evidence

Jan M. Ulijn and Judith B. Strother

Abstract—Writers must choose specific syntactic structures when writing 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 a significant factor in making syntactic choices? Four groups of students were tested: 48 American (L1) and 48 Dutch (L2)—half of whom had computer science backgrounds and half of whom did not. According to this experiment, neither ST knowledge nor language knowledge appeared to affect the choice of ST structures. Even without knowledge of the ST text's content matter, both L1 and L2 technical writers—experts significantly more than novices—tend to write in the ST register.

Introduction

A recent experiment (Strother & Ulijn, 1987) demonstrated that a thorough syntactic analysis is not required for reading comprehension and, therefore, syntactic rewriting (i.e., transforming ST [science and technology] syntax into CL [common language] syntax) does not affect EST (English for Science and Technology) comprehension. In writing on scientific or technical subjects, however, the writer cannot avoid choosing either CL or ST syntactic structuring. CL versions are characterized by content verbs, active constructions, and the use of clauses, whereas ST language versions are characterized by the use of nominalizations, participles, infinitives, and passive constructions.

The purpose of this study is to analyze differences in the syntactic structures used by groups of Dutch and American humanities and computer science university students when they were asked to complete sentences within a computer science text. It was predicted that computer science majors, especially those who are native writers of English, would use significantly more ST structures than humanities majors, especially those writing in English as a second language. Our data will be explained using current psycholinguistic theory of language production and the interlanguage concept. Also, our data may have applications to EST writing.

Background

Current approaches using protocol analysis (Flower & Hayes, 1981), online pause data (Matsuhashi & Quinn, 1984), and the signaled stopping technique

Address for correspondence: Jan M. Ulijn, Eindhoven University of Technology, Den Dolech 2, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
Judith B. Strother, Florida Institute of Technology, Language Institute, 150 University Boulevard, Melbourne, FL 32901, USA.
(Grunig, Ramsey, & Schneider, 1985) to observe the cognitive effects on writing do not have much to say about syntactic aspects of interlanguage in ST writing. Therefore, we draw upon a psycholinguistic theory of language production which was originally based on oral language (Kempen & Hoenkamp, forthcoming) and which has brought us to a minimal psycholinguistic model of writing which allows for three strategies: using serial, strictly parallel, or partially parallel processes for conceptualizing and formulating (transforming thoughts into a written text) (Ulijn 1987; Ulijn & Gobits 1986). In either strategy, syntax is an integral part of the formulation stage and writers cannot avoid choosing a particular syntactic structure. Previous research on Dutch L2 writers translating French–Dutch (F–D) contrastive structures from French ST instructions without the required technical background knowledge (Shadok Project, Ulijn & Kempen, 1976) and Dutch writers having to choose a syntactic structure for given words in CL instructions to find their way in an imaginary French town (Beausite Project, Ulijn, 1987) demonstrated that Dutch writers had more problems in translating or using F–D contrastive structures, whereas French L1 writers used F–D parallel (CL type) or F–D contrastive (ST–type) almost equally. In the experiment reported here, the question is: Do Dutch computer science students who tend to use the ST register in English differ from their American counterparts in syntax usage in an ST text and does their computer science knowledge affect this difference?

The next question pertains to how the interlanguage (IL) concept relates to the syntactic level in ST writing. The data presented by IL researchers such as Tarone (1985), Huebner (1985), Selinker & Gass (1984), and Selinker (1983) predominantly refer to syntactic aspects of oral languages which are rather unstable and variable across SL-users. Selinker (1984, and in this special English for Specific Purposes issue) asks questions such as:

1. Does ST context shape IL behavior?
2. Is complexification of IL different in ST contexts?
3. How does restricted special purpose language use differ for L1 and L2 writers?

What are the syntactic aspects of ST writing relevant to our questions so far? A number of researchers have analyzed the characteristics of English for Science and Technology (see Ulijn, 1984; Todd Trimble, Trimble, & Drobnic, 1978). These studies show that in ST text, some syntactic structures occur with greater frequency than they do in CL text. They include the use of the passive, nominalizations, and participle constructions.

Some researchers have concluded that the complexity of syntax in a particular text can cause difficulties with EST reading comprehension (see Foley, 1985; Mohammed & Swales, 1984; Charrow & Charrow, 1979; and Statman, 1976). For example, some problematic areas are thought to be "heavy noun phrase subjects and objects, syntactic markers of cohesion, and the role of non-technical vocabulary in technical texts. . . . Some other problematic areas that were found to cut across disciplines . . . include the interpretation of modals, the significance of punctuation or the lack of it, and various problems relating to long, complex sentences" (Cohen, Giasman, Rosenbaum-Cohen, Ferrara, & Fine, 1979, p. 556).
A recent experiment, however (Strother & Ulijn, 1987), demonstrated that a thorough syntactic analysis is not required for reading comprehension and that, therefore, syntactic rewriting (i.e., transforming ST syntax into CL syntax) does not affect EST comprehension. By implication, therefore, the reader can deal with whichever structures — either ST or CL — are present in a text.

The writer, however, must consciously choose either ST or CL syntactic structuring. It is assumed that most writers use the syntactic structures common to EST for the specific rhetorical functions of a scientific or technical subject. It has been postulated that there is a rhetorical hierarchy of EST (Lackstrom, Selinker, & Trimble, 1973):

- Level A — The Purpose of the Total Discourse
- Level B — The Functions of the Units that Develop the Purposes of Level A
- Level C — The Rhetorical Devices Employed to Develop the Functions of Level B
- Level D — The Relational Rhetorical Principles that Provide Cohesion Within the Units of Level C

The basic assumption of this hierarchy is that rhetorical function determines the grammatical choice of the writer (see also, Bley-Vroman, 1978).

A writer must be sure that the syntactic forms chosen are well-suited to express a specific meaning or idea. Widdowson (1977) calls this textualization, "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).

Selinker, Todd-Trimble & Trimble (1978) defined presupposition in EST discourse as "information that the writer assumes the reader shares with him." It has been concluded that "presuppositions shared by the technical writer and reader seem to affect surface syntax." (Lackstrom, Selinker, & Trimble, 1973). In a study such as this, would computer science majors—either Dutch or American — presuppose knowledge that humanities majors would not have at their disposal to use in writing within a computer science text? Would L1 writers be able to presuppose information that an L2 writer could not?

While it has been concluded that the subject of the text to be written — here a scientific or technical subject — determines the choice of syntactic structure to be used, no data was found in the literature on whether the background of the writer — scientific/technical versus nonscientific/nontechnical or L1 writer vs. L2 writer — is a strong influencing factor.

**Hypotheses**

If choice of syntax cannot be avoided in interlanguage and L1 ST writing (as opposed to reading), it is hypothesized that variation in the choice of syntactic structures should occur across the population in the following ways:
1. Writers with an expert knowledge, CS majors, will choose significantly more ST syntactic structures than novice writers in that field. Humanities majors, on the other hand, would not be able to master ST syntactic structures without knowledge of a relevant ST field.
2. Writers of English as a first language will choose significantly more ST syntactic structures than non-native writers because ST syntactic knowledge is part of total English knowledge.
3. Background knowledge contributes significantly more to L1 and interlanguage ST syntactic structuring in writing than language knowledge. This situation is unlike CL interlanguage and L1 language use where background knowledge is rather shared by a whole linguistic community.

Method

Subjects

Four groups of students were tested: 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 in The Netherlands. Half of each group were computer science students who were in their third or fourth year of study. The other half were humanities majors with no formal training in computer science. The Dutch students at Eindhoven 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 at the university with less exposure at the college (a high intermediate to advanced level).

Materials, Procedure, and Scoring

A slightly shortened version of a computer science article, that had no mathematical formulas or illustrations, was chosen for the experiment. The article,

| TABLE 1 | Number and Distribution of 96 Students Over the Experimental Conditions |
|-----------------|----------------|-----------------|-----------|
| BACKGROUND KNOWLEDGE | LANGUAGE KNOWLEDGE | TOTALS |
| | American (L_1) | Dutch (L_2) | |
| Computer Science Majors  | 24 | 24 | 48 |
| (Experts) | | | |
| Humanities Majors | 24 | 24 | 48 |
| (Novices) | | | |
| TOTALS | 48 | 48 | 96 |
"Mass Storage Systems and Evolution of Data Center Architecture," (Miller, 1982) came from the periodical Computer. Within the text, ten sentences were selected for the students to complete (Appendix B). In each of these sentences, it was possible to make either CL or ST syntactic choices. Students were given the words, in alphabetical order, needed to complete each sentence. The instructions told them they could change the order of the groups of words, change the forms of words (e.g., 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 the students understood the task, no specific feedback was given.

The following is an example of one of the sentences from the computer science text (see Appendix A for the first paragraph of the text):

(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.

Possible student solutions are:

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

(b) Remote terminals are connected by 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.

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

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

In this particular study, we determined the criteria for operationalizing the distinction between ST and CL syntactic structuring in the following way. Two independent raters were asked to evaluate the student solutions by analyzing the syntactic structure of the words between brackets and then to mark the solutions according to these directions:

1. If the 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.

In the examples given above, for the underlined fragments, (b), (c), and (d) are examples of S and (a) is an example of C. In this particular sentence, the interpretation of the verb connect is crucial. If the student tries to make a nominalization, a passive, or a participle of this verb, the score is S. If the student tries to conjugate the verb to make an active construction, the score is C.
Kies (1985) correctly states that stylistic variants between CL and ST registers are often due to functional differences, which explain, for instance, the use of active or passive. In the above case, however, the two solutions, S or C are equally possible in English, with hardly any difference in function.

Differences in opinion between the two raters were discussed and led to a definitive mark. Student answers were analyzed to decode logic problems. A logic problem occurred when the student 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 sentence J (see Appendix B), which should have read, "... the problem is separating and clustering sets of files and providing a mechanism...", a few students 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.

Design and Statistics

A 2 × 2 ANOVA analysis of variance (one way, with repeated measures) was used with the following independent variables and conditions (see Table 1):

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

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

The dependent variable was the number, ranging from 1 to 10, of ST syntactic versions of ten English sentences for each subject. A Mann-Whitney U-procedure (for more than 30 observations per factor) was applied to compute significant differences between the 2 × 2 conditions. In addition, the number of logic problems and noninterpretable syntactic structures, both of which could indicate noncomprehension of the text, was tabulated.
Results

Hypothesis 1 (writers with an expert knowledge—CS majors—will choose significantly more ST syntactic structures than novice writers in that field—humanities majors) and Hypothesis 3 (background knowledge contributes significantly more to L1 and interlanguage ST syntactic structuring in writing than language knowledge) are accepted. Hypothesis 2 (Writers of English as a first language will choose significantly more syntactic structures than L2 writers because ST syntactic knowledge is part of the general English knowledge) is rejected. Figure 1 shows the ST syntactic structure scores. For the sake of comparison, previous scores on a similar French–Dutch contrastive structure of an ST-type from French and Dutch students majoring in different academic fields are added. (See the above-mentioned Beausite Project, Ulijn, 1987).

The ANOVA analysis showed a significant overall effect due to the variance between the cells. $F(3,92) = 5.87$, $p < .005$. The Mann-Whitney U-procedure demonstrated significant differences between American CS and humanities
majors ($U = 2.04, p < .05$) and between Dutch CS and humanities majors ($V = 2.9, p < .005$) 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 ($U = 4.21, p < .005$) between the Dutch CS students and the American humanities students. The Dutch CS majors used the most ST syntactic structures (mean 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 fewer ST syntactic structures although not at a significant rate.

Further inspection of the data sentence by sentence demonstrates that just a few instances show the reverse effect: preference for the CL syntax rather than the ST syntax. All groups tended to avoid the passive voice in sentence A, whereas three of the groups avoided the nominalization in E and two, only the Americans, avoided the nominalization in G. Dutch humanities students tended to avoid the participle construction in I.

To what extent is this ST syntax score affected by noncomprehension of the ST text? The Dutch CS students produced the fewest logic problems, the most interpretable sentences, and the highest ST syntax score, although they had a lot of grammatical errors (for a further error analysis, see Strother & Uijn, forthcoming). The three other groups seemed to have more comprehension problems, although 142 incorrect responses out of 960 responses (14% noncomprehension) represent a very high comprehension rate.

**Discussion**

The results of this experiment will be discussed in light of the psycholinguistic theory of writing, the interlanguage approach, and specific aspects of ST writing.

**Psycholinguistic Aspects**

The strong priority for ST syntax use by L1 and L2 expert and novice writers, and the significant effect of background knowledge on ST syntax use (experts use more ST syntax than novices) present evidence that choice of syntactic structures cannot be avoided in writing. This finding contrasts with that of reading where syntax appeared to have hardly any effect. This finding, of course, does not say anything about the plausibility of serial, strictly parallel, or partially parallel strategies of conceptualizing and formulating.

The L2 writers used even more ST syntax than the L1 writers, although not significantly. The reason might be that the Dutch L2 writers make efforts to find stylistic support in the surrounding text because of their limited language knowledge, whereas the American L1 writers choose their own syntactic structure from the several options they have in their repertoire. For example, items B and G show this effect. In this particular context, the L2 syntactic knowledge is more focused toward ST.
Selinker has posed the following questions (1984, and in this special issue):

1. Does ST context shape IL behavior?
2. Is complexification of IL different in ST contexts?
3. How does restricted special purpose language use differ for L1 and L2 writers?

We present the following answers to those questions. If L1 writers did not use the ST register more than L2 writers, regardless of their knowledge of the subject matter of the text, one might conclude that ST interlanguage seems to be less marked syntactically in comparison with L1 language use than CL/ST linguistic descriptions suggest. Since background knowledge affects the ST syntax preference and language knowledge seems not to, the ST context appears to shape interlanguage behavior. If one considers ST syntax to be more complex, the IL strategy simplification has another meaning in ST writing by L2 experts: they use significantly more ST syntax than L1 novices. Therefore, complexification of IL in an ST context seems to be different.

In respect to the syntactic reduction phenomenon (Lantolf & Frawley, 1985), on the one hand, L2 writers use more ST syntax, which is more complex, because their English knowledge is more reduced to ST syntax. On the other hand, similar to oral language, the L2 writer uses an active instead of a passive construction in English because the active form is syntactically easier. (Item A, Appendix B). In addition, both L1 and L2 writers avoid some nominalizations (Items E and G).

Selinker (1984) also observed the following about a universal IL grammar: In specific purpose (ST) acquisition, language and subject matter are intimately and structurally intertwined. Our data show that ST syntax use appears to be influenced more by background knowledge than by language knowledge. Therefore, reduced special language use differs more for expert and novice writers than for L1 and L2 writers. Instability and variation have been pointed out to be characteristic interlanguage features referred to as transitional competence (Corder, 1971). In our data, L1 language use and interlanguage present the same syntactic variation or instability. This is in line with Tessman's (1985) findings that even L1 science writing is not stable in its style requirements. Questionnaires and interviews among researchers of eight national research laboratories in the U.S. show evidence that not all science writers avoid first person or insist upon using the passive voice. On the other hand, Bazerma (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.

So far there seems to be a striking similarity with ST reading. 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). Therefore, science readers probably just overlook all kinds of linguistic variation or instability in science writing. Science writers, on the other hand, deliberately allow for this variation or instability since their main attention is also triggered by purpose and schema in a constant interaction. Hence, it may not be wise to speak about interlanguage in L2 writing in science and technology since L1 writing in that area features the same syntactic instability or variation for the above reason.

On the other hand, it is necessary to warn against any overgeneralization of the apparent priority for ST syntax by all four test groups:

1. English might differ from other languages in this respect. In a similar experiment in CL writing by similar subjects (Ulijn, 1987), Dutch writers gave a strong preference to the Dutch–French (D–F) parallel syntax of a CL type, whereas French L1 writers balanced their choice almost perfectly between D–F parallel and D–F contrastive structures of an ST type. Moreover, the level of L2 knowledge might account for differences: Dutch students know English much better (high intermediate/advanced) than they know French (intermediate).

2. This test might not be sensitive enough to demonstrate differences in syntactic preference because of the high English knowledge level of the Dutch writers; that is, there is a ceiling effect.

3. The learning effect of the ST context on ST syntactic use might be avoided by a writing test with context-free sentences.

Topics such as these are indicative of areas of research that those concerned with interlanguage and ESP could address in future work.

Acknowledgments—We are grateful to the following for their assistance during the project. At Florida Institute of Technology, Dr. Edwin F. Strother, Prof. C. H. Gould, and Ms. M. F. Tucker helped with data collection and evaluation. At Eindhoven University of Technology, Dr. ir. Cees Memerik helped find the necessary subjects. Ms. Marga Engels, from the Maastricht School of Translation, as an intern, gave experimental assistance with Mr. Walter van Hulst, technology and communication student, who took care of the statistical analysis of the data.

(Received February 1987)

REFERENCES


**Jan M. Ulijn**, an associate professor in the Department of Applied Linguistics, Eindhoven University of Technology in The Netherlands, teaches Psycholinguistics & Technical Communicaton. He is the recipient of the *International

Judith B. Strother is Assistant Professor of Applied Linguistics and Assistant to the Director at Florida Institute of Technology's Language Institute in Melbourne, Florida. Her main interests are teaching and researching technical reading and writing. She has authored an EAP reading textbook for advanced ESL students entitled Kaleidoscope (Holt, Rinehart and Winston, 1986).

Appendix A

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 which 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 in paper, microfilm, or magnetic tape.

Appendix B

Directions: Using the following words and groups of words, write complete sentences.

A. (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.

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

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

E. (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.

F. Exactly what is meant by "getting the architecture straight" may be debatable; however, it means (stage) 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.

G. Certainly the I/O structure of the present-day computer does not constitute a clean interface, primarily because the operating systems assume (know) (the physical characteristics of the device) (. . .).

H. For example, most host computers do not permit (data file) (direct) (transfer) from magnetic tape to magnetic disk; they require reading the tape into main memory and then writing it out to the disk.

I. The term "backfill staging" has been used to indicate (direct) (the host) (. . .) (. . .) such a transfer without the movement of data into and out of the main memory.

J. When such movement is possible under system control, the problem is (and) (cluster) (separate) (sets of files) and (provide) a mechanism to have desired files more readily available.