

# Measuring the Translatability of Simplified English in Procedural Documents

Jan H. Spyridakis, *Member, IEEE*, Heather Holmback, and Serena K. Shubert

**Abstract**—This paper reports the results of a study that tested the translatability of a restricted language, called Simplified English (SE), as used in maintenance procedures in the airline industry. The study examined the effect of document type (SE versus non-SE) and procedure (procedure A versus procedure B) on the quality and ease of translation for native speakers of Spanish, Chinese, or Japanese. The results reveal that SE may be more effectively translated by native Spanish speakers than by Chinese speakers. The paper concludes with a discussion of methodological issues that researchers should consider when running such translation studies.

**Index Terms**—Airline industry, Chinese, English, Japanese, manuals, Simplified English, Spanish, translation.

THE use of a controlled or restricted subset of English for creating documentation is becoming more common in industry as companies attempt to improve the quality of technical documents, especially those that need to be read and understood by nonnative speakers of English. Restricted subsets of English are also being applied to documents that need to be translated into other languages, either by humans or computers. While the claims that using a restricted language makes documents more comprehensible and translatable are reasonable, empirical support is lacking.

In an attempt to provide empirical evidence as to whether, how, and to what degree a restricted language such as Simplified English (SE) can improve the comprehensibility and translatability of technical documentation, researchers from Boeing and the University of Washington conducted a two-part empirical study on SE, which is the writing standard currently used at Boeing for maintenance manuals. The first experiment, concerning the comprehensibility of SE, is reported in Shubert *et al.* [1]. Here we report on the second experiment, a pilot study on the effect of SE on the ease and accuracy of translation.

## OVERVIEW OF SE

Simplified English is one of several restricted language standards that have been developed to reduce ambiguity and provide greater consistency and readability in technical documents. Proponents of SE (and other controlled languages) have claimed that using a restricted English standard makes documents easier to read and understand, and easier to translate

accurately into other natural languages. SE was designed to be applied to both procedural and descriptive writing, but in practice it has primarily been applied to procedural technical documents, which we used in this experiment. What follows is a brief history and description of the SE standard. See Adriaens and Schreurs [2] for a more complete discussion of controlled languages.

The SE standard began to be formulated in the late 1970's. It was preceded by similar efforts, such as Charles Ogden's Basic English in 1932 [3] and Caterpillar Tractor Company's Caterpillar Fundamental English in 1972 [4], [5]. Caterpillar Fundamental English laid the foundation for the International Language for Servicing and Maintenance (ILSAM). ILSAM was developed to facilitate translation for international product support documentation [6]. SE was originally developed at Fokker, primarily by John Kirkman [7], and was officially adopted and modified by The Association Europeene de Constructeurs de Material Aerospatial (AECMA) for application to technical documentation in the aerospace industry [8]. SE is one of several descendants of Caterpillar's early work and ILSAM.

The SE standard was adopted by AECMA in response to requests from European airline companies to improve the readability of aircraft documentation. It is currently being used by almost all companies that produce aircraft maintenance procedures, including Aerospaciale Industrie, The Boeing Company, British Aerospace, Deutsche Aerospace, Fokker, General Electric, Lockheed, McDonnell Douglas, and Pratt & Whitney. Simplified English continues to be developed and refined by an international group, the AECMA SE Working Group, which released a new version (Issue 1) in late 1995 [9]. The Boeing Commercial Airplane Group (BCAG) has committed to comply with the SE standard in producing its maintenance manuals since it was adopted in the United States by the Air Transport Authority (ATA) and the Aircraft Industries Association (AIA). With the aid of an SE automated checker [8], the Customer Services Division of BCAG began distributing maintenance manuals written in SE in 1990.

The SE standard consists of a core vocabulary and a set of writing rules that govern grammar and style. There are also guidelines for company-defined technical vocabulary. The 1500-word core vocabulary consists of verbs, prepositions, conjunctions, adjectives, adverbs, and nouns. Words approved for the core vocabulary were chosen for their simplicity and commonality with other European languages [9]. In most cases, a given word is restricted to one meaning (to reduce lexical ambiguity), and a given meaning is represented in

Manuscript received June 1996; revised October 1996. This work was supported by The Boeing Company.

The authors are with the Department of Technical Communication, University of Washington, Seattle, WA 98195 USA, and the Boeing Co. (Holmback).  
 Publisher Item Identifier S 0361-1434(97)01825-0.

the vocabulary by only one word (to reduce synonymy). For example, “follow” can be used only in the meaning “to come after” and not in the meaning “obey”; and “start” is a legal SE word, but “begin” and “initiate” are not allowed.

Although the core vocabulary is limited, technical names and manufacturing processes can be added as needed. This flexibility ensures that the standard can be adapted to serve the different content domains of different companies. The goals of developing and using a highly restricted, standard vocabulary are to eliminate ambiguities and to choose the most common, simple word to express an idea to the reader. This rationale is supported by existing research on readability [10]–[15].

The writing rules of the SE standard were designed to simplify and make consistent the grammar and writing style of documents. The rules cover words, phrases, and sentences, as well as the form and content of procedures, the use of warnings and cautions, and the construction of paragraphs in descriptive writing. (See [9] for detailed SE writing rules.) In some ways, SE is very restrictive; it imposes constraints on 1) the length of sentences and noun compounds, 2) the types of verb forms that can be used (for example, no “-ing” forms, limited use of passive voice), 3) the organization of paragraphs (only one topic must be introduced in a paragraph, and it must be the first sentence of the paragraph), and 4) the expression of content (only one instruction per sentence). The “one word–one meaning” criterion can also be quite restrictive for proficient English writers, who have many synonyms at their disposal.

While it may be more difficult for writers to write in SE because of the restrictions, many companies use SE in producing documents, apparently operating under the assumption that SE accomplishes its goals of making documents easier to read, understand, and translate. Yet they seem to be doing so on the basis of intuitive and anecdotal evidence. Don Hinson raised his concerns about the “uncritical use of SE standards” in an article about issues for writers [6]. He states, “. . . AECMA’s SE claims to be founded on readability research. It would be interesting to establish the nature, validity, and appropriateness of the research used. It would also be helpful to know of any research carried out on SE manuals in use” (p. 36). Our recent experiment on the comprehensibility of SE supports the claim that using SE improves the comprehensibility of relatively complex technical aircraft maintenance manual procedures [1]. In the current experiment, we were interested in testing the assumption that SE improves the translatability of aircraft maintenance manual procedures for native speakers of different languages. We were also interested in reassessing how procedural complexity interacts with SE.

## METHODOLOGY

To test the effects of SE/non-SE, procedure, and native language on translation quality and ease, we had native speakers of either Spanish, Chinese, or Japanese translate one of four documents (SE or non-SE versions of Procedures A or B); then, raters from the same languages assessed the translations on several parameters. The specific methodology is discussed below; it is separated into the methods we used

for obtaining the translations and the methods we used for rating the translations.

### *Translation Methods*

*Subjects:* Subjects for the translation tasks consisted of native speakers of Chinese, Spanish, or Japanese. We used university students, rather than professional translators, as subjects for the following reasons: 1) we were interested in testing general claims about SE; 2) we believed that university students would represent novice translators and somewhat replicate the novice translator status of airline mechanics who sometimes translate portions of the maintenance procedures; and 3) we had limited funds for this pilot study.

To find subjects at the university to translate the documents, we ran advertisements in the University of Washington (UW) student newspaper. Many Chinese speakers responded and we selected those who had similar education levels (i.e., completed bachelor’s degrees). Many Spanish speakers responded to the ad and we secured more by word-of-mouth and through the Internet. The Spanish speaking subjects included UW undergraduates, two UW employees (one who had completed an undergraduate degree, and one who had not), and one housemate of a student. We used all of the few Japanese speaking respondents to the ad and secured more subjects by contacting UW Asian Student Services, ESL classes, primary engineering classes, Japanese student organizations, community colleges, religious and social organizations, and Internet sources. Nevertheless, we managed to recruit only six Japanese speaking undergraduate subjects for our experiment. Therefore, we did not statistically analyze the data for Japanese speakers as we did for the other two target languages. While we had planned to use 16 subjects per language, the final subject count included 18 native Chinese speakers, 15 native Spanish speakers, and 6 native Japanese speakers.

*Materials:* The translation materials consisted of four documents (two procedures written in SE and non-SE), an instruction sheet and consent form for each subject, and lined tablets or blank paper.

The translation documents consisted of the same two procedures employed in the earlier comprehension experiment [1]. The procedure for selecting these documents is discussed below. With the help of a Boeing Commercial Airplane Group (BCAG) technical editor, Paul Montague, a sample of fifteen non-SE naturally occurring maintenance manual procedures was collected. Each had an SE version, written by BCAG maintenance manual writers after compliance with the SE standard, and a non-SE version, written before compliance with the SE standard.

Because we wanted only two similar procedures for the final experiment, the sample had to be reduced. After reading all fifteen procedures to ascertain the range of procedures covered, team members did a rough count of word, sentence, and paragraph length to further narrow the sample. We initially identified five similar procedures in which the non-SE version fit the following criteria: no less than 450 words, no more than 1000 words, and no more than 15% passive voice.

The procedures were typed into Word 5.0®, and the team used the Word 5.0® grammar checker to further analyze the

documents. The information about number of words, number of paragraphs, percentage of passive voice, readability levels, and sentence length provided some useful criteria to help select the two final non-SE procedures, which we believed were similarly matched on the above parameters (see Table I). Furthermore, the number of steps were deemed similar in the two procedures and the topics in both were felt to be understandable by novices in the field. The final procedures chosen were Procedure A, entitled “49-15-01 APU Air Intake Duct-Removal/Installation,” and Procedure B, entitled “12-15-10” no. 1, 2 and 4 Passenger Door Emergency Power Reservoir—Servicing (Gaseous).” All documents (both SE and non-SE versions of the procedures) were reformatted to exhibit similar layouts.

While we believed these two procedures were sufficiently similar given the assessment criteria, the results of the comprehension experiment indicated otherwise: Procedure A was significantly more comprehensible when written in SE than non-SE, for both native and nonnative speakers of English, than Procedure B. A re-analysis of the two procedures led us to believe that Procedure A was more complex than Procedure B. Table I above reveals that Procedure A had fewer words and sentences, but more words per sentence and more paragraphs than Procedure B. And while Procedure A had fewer passive voice constructions than Procedure B, it had a higher Flesch rating. Boeing personnel also indicated that Procedure A involved a more complex task. Given these differences, we expected to see the procedure differences surface again when translation quality was measured.

*Procedure:* For the translation task, subjects attended one of many three-hour translation sessions (separate sessions were held for different languages). When subjects arrived at a session, the test administrator explained the general task from a pre-written script. Subjects were then randomly assigned to read and translate one of the four English source documents (subjects were not told anything about SE or non-SE). Written instructions told subjects to preserve the meaning and style of the original procedure when possible, to use and underline any English words they did not know, and to use any English words they believed should not be translated—yet not to underline them. Subjects completed a consent form and a brief demographic questionnaire asking number of years of education, number of years living in the US, and number of years of English instruction. Subjects were paid \$8 per hour for a maximum of three hours; most subjects used the entire three hours to complete the task.

#### Rating Methods

*Raters:* Three native speakers from each language (Chinese, Spanish, and Japanese) rated the translations. Raters were solicited through ads in the UW student newspaper and the Internet. The candidates submitted resumes, and the most qualified applicants were interviewed. Those chosen for the task demonstrated good communication skills in English (as exhibited in respondents’ cover letters, resumes, e-mail messages, and telephone conversations). All raters were UW graduate students.

TABLE I  
SELECTION CRITERIA FOR NON-SE PROCEDURES

Criteria	Procedure A	Procedure B
Counts		
Words	686	846
Sentences	32	63
Paragraphs	61	55
Averages		
Words per sentence	21	13
Characters per word	4	4
Readability		
Percent passive voice	3	11
Flesch grade level	14.7	8.8
Task Difficulty		
Stages	5	2
Steps	29	39

*Materials:* The rating materials included a baseline translation for each language; an English document with content ranked by importance; copies of the subjects’ translations; and a rating key.

There was one baseline translation for each document (i.e., the SE and non-SE versions of each procedure in each target language). These baseline translations were provided by three Boeing employees (one native speaker per language) who were familiar with maintenance manuals. The translations were designed to provide a standard for raters to consult and to exemplify “good” translations. Raters were also given a copy of the English documents ranked for content importance. Paul Montague, a BCAG technical editor, had ranked terms, sentences, and sections with a “1” for the more important information and a “2” for the less important information.

The rating keys included five measures, two of which had two submeasures: accuracy of the translation, style match with the original document, ease of comprehension, number of major and minor mistranslations, and number of major and minor omissions.

*Accuracy of the Translation:* Accuracy was rated on a five-point scale with 1 equal to “highly inaccurate” and 5 equal to “highly accurate.” Accuracy was defined as how precisely the translation reflected the content of the English source document.

*Style Match with the Original Document:* The degree to which the style of the translation matched the style of the original document was also rated on a five-point scale with 1 equal to “highly mismatches the original” and 5 equal to “highly matches the original.” Style was defined as the language level, sentence difficulty, and tone of the document. This measure was based on the assumption that if a document was easy to translate, the style of the original document would be preserved in the translation, as subjects had been instructed to maintain the style of the original document when possible.

*Ease of Comprehension:* Ease of comprehension was also rated on a five-point scale with 1 equal to “very unclear” and 5 equal to “very clear.” Raters were instructed to rate the translations for how clear and easy they were to read and

comprehend. This measure was based on the rationale that a more readable translation results when the translation task is relatively easy.

*Number of Major and Minor Mistranslations:* The number of mistranslations and the number of omissions are traditional measures of translation quality. Mistranslations were defined as words that are inaccurately translated. Raters identified and ranked mistranslated words as either major or minor by comparing them with the baseline translations' terms, sentences, and sections ranked for importance. Major mistranslations corresponded to a 1 and minor mistranslations to a 2 on the baseline translations.

*Number of Major and Minor Omissions:* Omissions were defined as words that subjects left out of their translations. Raters identified and ranked omissions as either major or minor by comparing the translations with the baseline translations' terms, sentences, and sections ranked for importance. Major omissions corresponded to a 1 and minor omissions to a 2 on the baseline translations. Raters were instructed to ignore the omission of function words (such as articles). Further, the raters were trained not to assess terms that should not be translated as omissions (e.g., label on a part). The baseline translations helped ensure interrater reliability regarding what terms should or should not be translated.

*Procedure:* Raters were trained at an explanatory meeting and a review session (separate meetings were held with raters of the three languages). At the explanatory meeting, the administrator explained the task and provided detailed instructions. Raters read the baseline translations carefully and agreed on terms that they thought should be translated and altered the baseline accordingly (e.g., many terms were not translated in the original Spanish baseline). Further, raters gained an understanding of and agreed on how the measures were to be operationalized. After the training meeting, raters were given a practice translation to rate before the review session. The practice translation (one per language) was randomly selected from the translations completed during the experiment (the ratings of these practice translations were excluded from the analysis of the experimental data). The ratings of these translations were discussed at the review meeting and indicated that sufficient agreement among the raters had been established. The rater agreement was reconfirmed in the actual experiment. After the review meeting, raters were given copies of all of the translations for their target language. The raters were paid \$240.00 each for their efforts.

After the raters returned the rated translations, the ratings of the three raters per language were averaged to provide one rating score per dependent measure for each subject. These scores were analyzed in Statview 4.0.

## RESULTS AND DISCUSSION

The primary goal of this pilot study was to investigate the claim that the use of SE in a technical procedure will improve the translatability (i.e., quality and ease of translation) of that document compared with the same technical procedure not written in SE. Another goal of the study was to assess whether there were differences in translation quality and ease

between the two procedures (A and B), as we were interested in how document complexity, a factor that arose in our earlier comprehension study, might influence translatability. We were also interested in any differences in translation quality among the different target languages (Chinese, Spanish, Japanese) to determine whether SE affects translation quality equally for the three languages. To assess the issues described, we wanted to run a three-way ANOVA (SE/Non-SE  $\times$  Procedure  $\times$  Language); however, the small cell sizes for the Japanese speakers precluded using them in such an analysis. Instead, three  $2 \times 2$  ANOVAs (SE/non-SE  $\times$  Procedure) were run on the seven dependent measures for 1) translations from all languages combined, 2) Spanish translations, and 3) Chinese translations. Although we statistically analyzed the Spanish and Chinese translations, we were concerned about the small cell sizes and the resulting potential for large variances that might make it difficult to obtain significant findings.

Because the study reported here was difficult to conduct because of very little precedent in the literature for measuring ease and quality of translation, another goal of this study was to assess the design of this experiment and identify areas for further exploration. Because of this and the small cell sizes, we report patterns that appeared in the data as well as the results of inferential statistics.

After a brief discussion of the demographic data, the main effects of SE/non-SE and Procedure are discussed, first for all languages combined and then individually for Spanish and Chinese. This discussion is followed by a discussion of interactions.

### *Relationship Between Translations and Subject Demographics*

The demographic profile of subjects consisted of three variables: number of years of education, number of years living in the US, and number of years of English instruction. From this data, we were attempting to identify a measure of English ability, as TOEFL scores would have been either unavailable or extremely outdated for many subjects.

To determine whether the speakers of the three languages differed, we conducted three one-way ANOVAs of native language on the three demographic variables. The ANOVAs were nonsignificant, suggesting that the speakers of the three languages were similar in number of years of education, number of years living in the U.S., and number of years of English instruction. We then correlated the demographic data with the translation ratings and found no consistent patterns. We have concluded that one may need to have a more direct measure of English ability to see how it might affect the translation quality of SE versus non-SE documents. Subjects' level of literacy in their native language and experience with scientific or technical material might also be relevant factors.

### *Effect of SE Versus Non-SE*

We first examine the results for the effect of document type (SE versus non-SE) on the accuracy, style match, comprehension, mistranslations (minor and major), and omissions (minor and major) for the target translations.

*Languages Combined:* For the three languages combined, a significant main effect ( $p < 0.05$ ) occurred only for the style match and the minor omissions measures. In both cases, the SE translations were superior, matching the style of the original document better and containing fewer minor omissions than the non-SE translations. Table II lists the means and standard deviations (in parentheses) for both the SE and non-SE documents, and the  $F$  and  $p$  values associated with the ANOVAs for the seven dependent measures.

Although significant differences were limited for all languages combined, the data do reveal an interesting pattern: subjects who translated SE documents produced higher quality translations than those who translated non-SE documents (except for major mistranslations). While these differences were significant on only two measures, the large variance on some measures, caused perhaps by combining the languages, may account for the lack of significance on the other measures. Hence, an analysis of the languages separately may be more revealing.

*Spanish Translations:* For the Spanish speakers, translations of the SE documents differed significantly from the non-SE documents on four measures: SE translations scored significantly higher on accuracy, style match, and comprehensibility, and contained significantly fewer minor mistranslations than non-SE translations. Table III lists the means and standard deviations (in parentheses) for both the SE and non-SE documents, and the  $F$  and  $p$  values associated with the ANOVAs for the seven dependent measures.

Beyond the significant differences between SE and non-SE means on four measures, for five measures (accuracy, style match, comprehension, minor mistranslations, and major omissions) the variances are significantly heterogeneous ( $p \leq 0.05$ ) between the SE and non-SE translations. And the variance on these measures is consistently smaller for the SE translations than for the non-SE translations, suggesting that the SE translations are more consistent than the non-SE translations. A similar pattern is seen for the other two dependent measures, although the variances do not differ significantly. Fortunately, the heterogeneity between SE and non-SE variances is not so extreme as to invalidate the ANOVA statistics [16].

In summary, for the Spanish speakers, SE translations received significantly better ratings on four measures and showed significantly greater consistency on five measures than the non-SE translations.

*Chinese Translations:* The ANOVA results as well as the means and standard deviations (in parentheses) for Chinese speakers translating the SE or non-SE documents are presented in Table IV. There were no significant differences between the SE and non-SE translations. Further, the variances associated with the means are generally homogeneous, suggesting that the SE and non-SE documents were translated with equal consistency.

*Comparison of Spanish and Chinese Translation Results:* For both the Spanish and Chinese translations, an examination of both the significant and nonsignificant differences reveals that the SE translations were rated better than the non-SE

TABLE II  
SE VERSUS NON-SE DOCUMENTS FOR LANGUAGES COMBINED

Measures	Means (SD)		$F$	$p$
	SE n = 19	non-SE n = 18		
Accuracy	3.72 (1.19)	3.11 (1.32)	2.15	.1521
Style Match	4.01 (.91)	3.41 (1.16)	4.74	.0367*
Comprehension	4.00 (1.20)	3.28 (1.45)	2.75	.1069
Mistranslations				
Major	10.02 (11.55)	9.37 (11.48)	0.005	.9464
Minor	7.46 (5.75)	10.57 (7.05)	2.112	.1559
Omissions				
Major	4.35 (5.86)	6.33 (10.52)	0.751	.3927
Minor	2.19 (2.62)	9.67 (11.93)	7.63	.0093*

\* indicates  $p \leq .05$

TABLE III  
SPANISH TRANSLATIONS OF SE VERSUS NON-SE DOCUMENTS

Measures	Means (SD)		$F$	$p$
	SE n = 8	non-SE n = 6		
Accuracy	4.50 (.31)	3.56 (1.15)	5.14	.0467*
Style Match	4.71 (.28)	3.61 (.74)	23.07	.0007*
Comprehension	4.75 (.29)	3.44 (1.40)	5.56	.0400*
Mistranslations				
Major	4.38 (2.17)	4.83 (4.00)	0.09	.7617
Minor	1.17 (2.12)	10.33 (5.87)	12.79	.0050*
Omissions				
Major	2.58 (2.21)	1.13 (1.33)	1.70	.2241
Minor	1.17 (.87)	13.22 (15.31)	3.28	.1001

translations on twelve of the fourteen measures: higher scores for accuracy, style match, and comprehension; lower scores for both types of mistranslations and omissions. The two exceptions are for Spanish major omissions and Chinese major mistranslations.

There were more significant differences, and larger differences between means even when there was no significance, for the Spanish SE translations versus the non-SE translations than for the Chinese translations, or for the combined languages' data (which included the Japanese subjects). One reason for

TABLE IV  
CHINESE TRANSLATIONS OF SE VERSUS NON-SE DOCUMENTS

Measures	Means (SD)		F	p
	SE n = 9	non-SE n = 8		
Accuracy	3.00 (1.32)	2.54 (1.51)	.51	.4843
Style Match	3.67 (.97)	3.25 (1.46)	.75	.3997
Comprehension	3.11 (1.20)	2.88 (1.55)	.15	.7034
Mistranslations				
Major	17.07 (13.68)	16.63 (14.02)	.02	.8753
Minor	11.78 (5.47)	13.96 (6.64)	.61	.4486
Omissions				
Major	6.78 (7.71)	11.79 (12.95)	1.26	.2816
Minor	3.33 (3.46)	10.13 (10.79)	3.67	.0774

this might be that English and Spanish are linguistically more similar and therefore the benefits of SE could be more obviously transferred to translations by Spanish speakers. This similarity should improve the quality of the resulting translations, and it should also have some positive effect on the ease of translation.

A finding that relates to our interpretation of the effect of the similarity of Spanish to English occurred in a study of native speakers of Arabic, Spanish, "Oriental," and "Other" recalling English texts written in different rhetorical forms [17]. Carrell found that the native speakers who possessed the appropriate schema to process the various discourse types did best on the recall test—and in her study the Spanish speaking subjects performed better than the speakers of the non-European languages.

While significantly better Chinese translations of the SE documents would have further supported our hypothesis that SE helps the quality and ease of translation, the fact that Chinese is not as linguistically similar to English as Spanish may have several explanations. The Chinese speakers may be less sensitive linguistically to the differences between the SE and non-SE, or perhaps, the English structures are changed so much in the process of translating the document into Chinese (as opposed to Spanish) that the SE/non-SE differences are not reflected in the Chinese translations. To better measure the effects of SE versus non-SE source documents on the process of translating from English into another languages, future studies will need to consider using speakers of many other languages as subjects (European and non-European) as well as perhaps assessing additional measures of quality and ease of translation.

#### *Effect of Procedure*

Beyond differences between SE and non-SE source documents, the ANOVAs assessed differences between the pro-

cedures. In our comprehension experiment, subjects reading the SE version of Procedure A performed significantly better on comprehension measures than subjects reading the non-SE version; on Procedure B, the difference was not significant. As discussed earlier, we concluded that the non-SE version of Procedure A was more complex than the non-SE version of Procedure B [1]. In analyzing the results of the current study, we were interested in assessing whether such document differences would occur in the translation results.

For all languages combined, the only significant main effect for procedure occurred for major mistranslations,  $F(1, 33) = 6.635$ ,  $p = 0.02$ , with Procedure A containing fewer major mistranslations ( $M = 4.7$ ;  $SD = 2.84$ ) than Procedure B ( $M = 13.93$ ;  $SD = 14.03$ ). No main effects for procedure were found for the Spanish translations. For the Chinese translations, again there was a main effect for procedure for major mistranslations,  $F(1, 13) = 16.92$ ,  $p = 0.0012$ , with Procedure A containing fewer major mistranslations ( $M = 6.5$ ;  $SD = 2.11$ ) than Procedure B ( $M = 26.07$ ;  $SD = 14.03$ ). This result, which replicates the result for all languages combined, clearly triggered the significant effect for all languages combined. Although we were surprised to find virtually no main effects for procedures, an examination of the interactions reveals some procedure differences.

#### *Interactions Between SE/Non-SE and Procedure*

This last examination of the results concerns the interactions between SE and non-SE documents for the two different procedures.

*Languages Combined:* For all languages combined, the only significant interaction occurred for minor omissions,  $F(1, 32) = 4.70$ ,  $p = 0.04$ . As Table V reveals, the number of minor omissions for the non-SE translations of Procedure A was significantly higher ( $M = 14.15$ ;  $SD = 15.04$ ) than Procedure A's SE translations ( $M = 1.04$ ;  $SD = 0.84$ ), Procedure B's non-SE translations ( $M = 4.63$ ;  $SD = 3.43$ ), and Procedure B's SE translations ( $M = 3.03$ ;  $SD = 3.16$ ).

An examination of Table V reveals some interesting patterns. First, virtually all SE translations received better ratings than non-SE translations (see  $\wedge$  on Table V for the two exceptions). However, as Table V illustrates, the differences in favor of the SE translations versus the non-SE translations were generally larger for Procedure A, the document that the comprehensibility study deemed to be more complex, than for Procedure B. While these differences were not large enough to result in numerous significant interactions, perhaps because of small cell sizes, the pattern of potential interactions in favor of the SE translations of Procedure A exists. This pattern replicates the pattern found in the comprehension study.

While much of the pattern in the means in Table V supports Procedure A as the more complex document, some does not. What is unclear from this small study is whether a source document's complexity might affect comprehensibility but not translatability or whether larger cell sizes might have made the translation results reveal more significant interactions and fully replicate the comprehension study's results.

TABLE V  
MEANS FOR LANGUAGES COMBINED

Measures	Means			
	Procedure A		Procedure B	
	SE n = 8	non-SE n = 9	SE n = 11	non-SE n = 9
Accuracy	3.83	3.00	3.64	3.22
Style Match	4.54	3.41	3.76	3.41
Comprehension	4.17	3.38	3.88	3.19
Mistranslations				
Major	3.50	5.82	14.76 <sup>^</sup>	12.93
Minor	6.54 <sup>^</sup>	5.82	6.54	11.22
Omissions				
Major	1.08	4.75	6.73	7.92
Minor	1.04	14.15	3.03	4.63

TABLE VI  
MEANS FOR SPANISH TRANSLATIONS

Measures	Means			
	Procedure A		Procedure B	
	SE n = 4	non-SE n = 4	SE n = 4	non-SE n = 2
Accuracy	4.42	3.75	4.58	3.17
Style Match	4.75	3.91	4.67	3.00
Comprehension	4.67	3.50	4.83	3.33
Mistranslations				
Major	2.41	4.75	6.33 <sup>^</sup>	5.00
Minor	2.67	8.25	5.17	14.50
Omissions				
Major	1.42 <sup>^</sup>	0.78	3.75 <sup>^</sup>	1.67
Minor	1.00	17.42	1.33	4.83

*Spanish Translations:* There were no significant interactions for the Spanish translations. The means for all conditions of the Spanish translations are shown in Table VI. The patterns in the SE/non-SE data for both procedures replicate the means for the main effects shown in Table III: the translations of the SE documents were rated better for both procedures A and B, with only three exceptions (see <sup>^</sup> in Table VI).

*Chinese Translations:* There were no significant interactions for the Chinese translations. Table VII lists the means for all conditions. The translations from the SE documents were rated better (two exceptions denoted by <sup>^</sup> in Table VII) than the translations from the non-SE documents; however, Procedure A exhibits larger differences between the translations of the SE versus non-SE documents than does Procedure B.

The greater positive effect of SE versus non-SE for Procedure A compared to Procedure B in the Chinese data was probably responsible for the similar pattern noted in the data for all languages combined. Based on this pattern in the Chinese data and the similar results of our earlier comprehension study, we would expect that using a source document written in SE would produce relatively better trans-

TABLE VII  
MEANS FOR CHINESE TRANSLATIONS

Measures	Means			
	Procedure A		Procedure B	
	SE (n = 4)	non-SE (n = 4)	SE (n = 5)	non-SE (n = 4)
Accuracy	3.50	2.25	2.60	2.83
Style Match	4.42	2.92	3.07	3.56
Comprehension	3.50	2.92	2.80	2.83
Mistranslations				
Major	5.00	8.00	26.73 <sup>^</sup>	25.25
Minor	10.33	16.42	12.93 <sup>^</sup>	11.50
Omissions				
Major	1.17	8.75	8.75	14.83
Minor	1.00	14.25	5.20	6.00

lations for Procedure A than for Procedure B (assuming that more complex documents can benefit most from the goals or SE). Questions remain, however, as to why this pattern is not significant and why it is not observed in the Spanish translations. The answer to the first question may relate to the small cell sizes. To answer the second question, one might speculate that the overall translation task was easier for the Spanish subjects (because of the linguistic similarities between English and Spanish) so it might have been easier for the translators to process and more adequately translate even the more difficult documents in the time allotted and hence exhibit fewer differences between the two procedures.

#### CONCLUSIONS AND FUTURE RESEARCH

While most of our conclusions are fairly speculative, given the exploratory nature of our study, it does appear that in certain cases using SE as a source language improves the quality and ease of translations. For the Spanish speakers, SE translations received significantly better ratings on four measures and showed significantly greater consistency on five measures than the non-SE translations. The occurrence of these findings for the Spanish translations but not for the Chinese translations may be due to the greater linguistic similarity of Spanish to English than Chinese to English. It may well be that SE will prove to be of greater benefit to translators whose native languages are linguistically more similar to English, though certainly more work is needed to reach such a conclusion with any certainty.

At this point, conclusions regarding how SE is affected by document complexity are hard to draw. In the qualitative analysis of the data, a pattern arose in the Chinese translations, showing a greater positive effect of SE versus non-SE for the more complex Procedure A compared to less complex Procedure B. It is unclear as to why such a pattern was not strong enough to be significant (though our sample was small) and why the pattern did not occur for Spanish speakers. Again, the answer may lie in the closer tie between Spanish and English; perhaps for translators whose native language is linguistically more similar to English, document complexity (as represented

by Procedure A) was not sufficiently problematic to trigger the purported benefits of SE. Certainly more work is needed to solve this dilemma.

As stated earlier, another goal of this study was to assess the design of this experiment and identify areas for further exploration. Because there is very little precedent in the literature for this type of work, we feel that this study will help contribute to future studies, not only in its results, but in what it suggests about the design of translatability studies in general.

In assessing the translations, the researchers, the subjects, the raters, and the Boeing employees who provided the target translations were concerned about: 1) which words were best left untranslated in a technical document, and 2) how closely a translation should replicate the original document to be considered a good translation. In this study, with the goal of interrater reliability, these issues were decided among the raters. Another concern regarded the raters' and subjects' unspoken value judgments about simple language and how such judgments might affect the results. Speculating on how to handle these issues as a general rule for all translation studies was beyond the scope of the current experiment, though we recognize that they need to be better addressed if translation studies are to become comparable and lead to theory development. And, of course, these issues are important, not only to SE, but to the field of translation as a whole.

Future research will also need to focus on what to measure and how to analyze translations. In future studies, we would recommend a refinement of measures to more directly test both quality and, especially, ease of translation. Our greatest dilemma concerned how to measure ease of translation. We used the measure of "style match" to indirectly assess the ease of translation with the belief that if a document were easy to translate, the style of the original document would be preserved in the translation. We also reasoned that more comprehensible translations might result when the translation task was easier. Although we asked subjects how easy the task was, we did not get many usable responses. We might additionally recommend that linguistic analysis be performed to provide a better understanding of how the translations compare to the originals, e.g., one could examine the mistranslations and look for systematic causes stemming from the original documents. Such an analysis, however, was beyond the scope of the current exploratory study.

Future studies should also consider the number and type of subjects and documents to use. To make better use of inferential statistics in a study such as this one, more subjects would need to be recruited. In enlarging the total sample size, one could also seek to refine the measures of language literacy. While the means for years of education, years of English, and years in the U.S. did not significantly differ among our subjects, with a larger sample one could assess other measures of English ability as well as measures of native language literacy and experience with technical documents, to see if they interact with the effect of SE versus non-SE and document complexity. Researchers might consider using subjects who are professional translators of many different European and non-European languages, yet we must remember

that, in the real world, nonprofessional translators do translate documents. Finally, future studies will have to use more and more documents, differing in length, difficulty, purpose, and so on. It is only by considering the issues discussed here that we can advance the research on the translatability of SE and further our understanding of when and whether SE contributes to the translatability of technical documents.

While our study was primarily exploratory in nature, we can suggest some ways in which a company whose documents are frequently translated might use our results. Such a company should definitely investigate using SE or a similar controlled language as a standard for authoring procedural documents that are to be translated, particularly when translations are from English to other Indo-European languages and when non-professional translators are employed. A company might also want to conduct its own study to assess the translatability of its documents and the benefits of using a controlled language in its specific setting. The design, results, and issues raised in the current study should contribute to such an undertaking.

#### ACKNOWLEDGMENT

The authors wish to thank Jim Hoard, Rick Wojcik, and Oscar Kipersztok, Boeing Information & Support Services, and Paul Montague, Tony H. Kuo, and Frank Tou, BCAG Customer Services Division, for their assistance with this project and for comments on an earlier version of this paper.

#### REFERENCES

- [1] S. K. Shubert, J. H. Spyridakis, H. K. Holmback, and M. B. Coney, "The comprehensibility of Simplified English in procedures," *J. Tech. Writing and Commun.*, vol. 25, no. 4, pp. 347-369, 1995.
- [2] G. Adriaens and D. Schreurs, "From cogram to algorithm: Toward a controlled English grammar checker," in *Proc. COLING-92*, 1992, pp. 595-601.
- [3] C. K. Ogden, *Basic English, A General Introduction with Rules and Grammar*. London, England: Paul Trebor & Co., 1932.
- [4] Caterpillar Corporation, *Dictionary for Caterpillar Fundamental English*. East Peoria, IL: Caterpillar Corp., 1974.
- [5] C. A. Verbeke, "Caterpillar Fundamental English," *Training and Devel. J.*, vol. 27, no. 2, pp. 36-40, 1973.
- [6] D. E. Hinson, "Simplified English—Is it really simple?" *WE*, pp. 33-36, 1988.
- [7] J. Kirkman, "How to communicate computerese," *Data Syst.*, pp. 14-15, July 1978.
- [8] J. E. Hoard, R. Wojcik, and K. Holzhauser, "An automated grammar and style checker for writers of Simplified English," in *Computers and Writing, State of the Art*. Oxford, England: Intellect, 1992, pp. 278-296.
- [9] AECMA, "AECMA Simplified English Standard," AECMA Doc. PSC-85-16598, Issue 1, Brussels, Belgium, 1995.
- [10] P. T. W. Hudson and M. W. Bergman, "Lexical knowledge in word recognition: Word length and word frequency in naming and decision tasks," *J. Memory and Language*, vol. 24, pp. 46-58, 1985.
- [11] M. Naveh-Benjamin and T. Ayres, "Digit span, reading rate, and linguistic relativity," *Quart. J. Exper. Psychol.*, vol. 38A, pp. 739-751, 1986.
- [12] J. G. Kroll and J. S. Merves, "Lexical access for concrete and abstract words," *J. Exper. Psychol.*, vol. 12, no. 1, pp. 92-107, 1986.
- [13] M. J. Lewellen, S. D. Goldinger, D. B. Pisoni, and B. G. Greene, "Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization," *J. Exper. Psychol.*, vol. 122, no. 3, pp. 316-330, 1993.
- [14] A.R. Dobbs, A. Friedman, and J. Lloyd, "Frequency effects in lexical decisions: A test of the verification model," *J. Exper. Psychol.: Human Perception and Performance*, vol. 11, no. 1, pp. 81-92, 1985.
- [15] G. B. Simpson and C. Burgess, "Activation and selection processes in the recognition of ambiguous words," *J. Exper. Psychol.*, vol. 11, no. 1, pp. 28-39, 1985.

- [16] D. C. Howell, *Statistical Methods for Psychology*. Boston, MA: Duxbury Press, 1982.
- [17] P. L. Carrell, "The effects of rhetorical organization on ESL readers," *TESOL Quart.*, vol. 18, no. 3, pp. 441–469, 1984.
- Heather Holmback** has a Ph.D. degree in linguistics and since 1986 has been involved in computational linguistics. She has worked on general natural language processing systems and specific applications, concentrating on semantic and discourse processing. Her most recent work at Boeing includes leading a project to develop meaning-based language checking applications.

**Jan H. Spyridakis** is an Associate Professor in the Department of Technical Communication at the University of Washington, Seattle. Her recent work on the comprehensibility and translatability of Simplified English for Boeing BCAG is an extension of her work in refining research methods in technical communication and in assessing document and screen design variables that affect usability and comprehension.

**Serena K. Shubert** is pursuing doctoral studies in Educational Curriculum and Instruction at the University of Washington, Seattle, where she is continuing her research on controlled languages. Her Master's thesis in technical communication concerns the comprehensibility of Simplified English.