
Text simplification and comprehensible input: A case for an intuitive approach

Language Teaching Research

0(0) 1–19

© The Author(s) 2012

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1362168811423456

ltr.sagepub.com

ltr.sagepub.com



Scott A. Crossley

Georgia State University, USA

David Allen

The University of Tokyo, Japan

Danielle S. McNamara

[The University of Memphis](http://www.memphis.edu), USA

Abstract

Texts are routinely simplified to make them more comprehensible for second language learners. However, the effects of simplification upon the linguistic features of texts remain largely unexplored. Here we examine the effects of one type of text simplification: intuitive text simplification. We use the computational tool, Coh-Metrix, to examine linguistic differences between proficiency levels of a corpus of 300 news texts that had been simplified to three levels of simplification (beginner, intermediate, advanced). The main analysis reveals significant differences between levels for a wide range of linguistic features, particularly between beginner and advanced levels. The results show that lower level texts are generally less lexically and syntactically sophisticated than higher-level texts. The analysis also reveals that lower level texts contain more cohesive features than higher-level texts. The analysis also provides strong evidence that these linguistic features can be used to classify levels of simplified reading texts. Overall, the findings support the notion that intuitively simplified texts at the beginning level contain more linguistic features related to comprehensible input than intuitively simplified texts at the advanced level.

Keywords

text simplification, intuitive simplification, corpus linguistics, computational linguistics, text comprehensibility

Corresponding author:

Scott A. Crossley, Georgia State University, Department of Applied Linguistics/ESL, 34 Peachtree Street, Suite 1200, One Park Tower Building, Atlanta, GA 30303, USA

Email: sacrossley@gmail.com

1 Introduction

The role of input is critical in second language acquisition (Hatch, 1978). The input that second language (L2) learners receive is often modified to make it more comprehensible. In spoken speech, L2 learners receive modified input that is simplified at the lexical, phonological, and syntactic level (Gaies, 1983). At the level of written texts, input is often modified to make the text more comprehensible, generally in the syntactic structure and the lexicon (Hill, 1997). When selecting reading material, material developers and teachers have two other choices as well: elaborated texts and authentic texts. Authentic texts are texts that were originally created to fulfill a social purpose in a first language community (Little et al., 1989). Elaborated texts maintain the complexity of the authentic text, but clarify message content and structure through repetition and paraphrasing (Yano et al., 1994).

Generalizations about comprehension effects that result from authentic, simplified, and elaborated texts are difficult because of differences in research designs; however, research does show that simplification generally improve literal comprehension. Overall, text modification of some sort tends to improve comprehension (Yano et al., 1994). For this reason, modified texts are more common in the L2 classroom than authentic texts (Young, 1999).

While structural processes of simplification (i.e. the use of word lists or readability formulas) may seem clear, the effects of text modification are far from clear. That is to say, texts are generally simplified to make them more comprehensible by modifying their lexicon and their syntactic structure (Simensen, 1987; Young, 1999). However, the process of simplification may have unintended consequences and, in some cases, potentially produce texts that are less comprehensible. For instance the dependence on high frequency vocabulary may unintentionally lead to the use of more ambiguous words. Likewise, the simplification of vocabulary may lead to a condensed syntactic structure that is more complex (Crossley et al., 2007). This may be especially true when authors depend on their intuition when simplifying text.

The purpose of this article is to investigate the effects of intuitive text simplification when topic and level are controlled. Thus, we examine the effects of simplification on 100 authentic texts simplified to beginning, intermediate, and advanced levels. We distinguish differences between the text levels using data taken from the computational tool Coh-Metrix (Graesser et al., 2004) and extend the findings to an analysis of text cohesion and text sophistication. Our goal is to examine the benefits or disadvantages of intuitive text simplification across proficiency levels. We hypothesize that intuitive text simplification will produce simplified texts that contain more linguistic features related to text comprehension and readability as the level decreases because the simplification task depends on human discourse processing, which links directly to text processing and comprehension. We also investigate the potential for unintentional changes in text that may result from intuitive text simplification.

1 Text simplification

Texts simplified for L2 readers come in a variety of forms. They can be adapted or abridged versions of original texts, or texts written for the teaching of specific grammar

or linguistic forms. Recently, there has been a call for more texts written directly for the L2 target audience that are more natural, but still control for lexical and grammatical content (Hill, 2008). However, all simplified texts share the same goal: increased comprehensibility and reduced cognitive load. The primary methods of attaining this comprehensibility, as discussed earlier, are in the modification of the lexicon and the syntax. In addition, publishers also consider the subject matter of the text, the cultural and background knowledge needed to understand the text, the learner, and the literary merit of the text (Lotherington-Woloszyn, 1993). However, for the purposes of this article, we are only interested in linguistic modifications.

Linguistically, L2 text writers have two choices when simplifying a text for an L2 reading audience: a structural approach or an intuitive approach (Allen, 2009). The structural approach to simplification is commonly found in graded reader schemes, which are aimed at advancing learners' language acquisition through extensive reading. In graded readers, authors utilize word lists and structure lists that are predefined for each level of the series (Nunan, 1999). Another approach subsumed under the structural approach is text simplification guided by the use of traditional readability formulas. Traditional readability formulas are simple algorithms that measure text readability based on sentence length and word length. They have found success in predicting L1 text readability, but have been widely criticized by discourse analysts (Davison & Kantor, 1982) as being weak indicators of comprehensibility. The use of readability formulas in guiding L2 text simplification has also been heavily criticized (Carrell, 1987), and readability formulas have been demonstrated to be less effective at predicting text difficulty than formulas derived from indices that tap into cognitive processing (e.g. decoding, syntactic processing, and meaning construction; Crossley et al., 2008). Regardless, the use of traditional readability formulas to develop simplified texts is quite common (Greenfield, 2004).

Intuitive approaches to L2 text simplification are also common. Even with recourse to word lists and structure lists for reference, authors following a structural approach still report relying mainly on their intuition (Young, 1999). Author intuition is influenced by personal beliefs (Lotherington-Woloszyn, 1993) and simple hunches about what makes a text more readable. Under an intuitive approach, the author's experiences as a language teacher, language learner, or materials writer (or any combination of these) guide the process of simplification and allow the authors to rely on their own subjective approximations of what learners at a particular level should be able to understand (Allen, 2009). While it is not known how common an intuitive approach toward simplification is when compared to a structural approach, Simensen (1987) found that while many publishers offered advice on adapting texts, most writers were still likely to depend on intuition. Simensen's research along with that of Young (1999) and others provide evidence that an intuitive approach might not only be extremely common, but perhaps the most common strategy in L2 text simplification.

Regardless of how a text is simplified, the end result of simplification and its perceived benefits to L2 learners is of considerable interest to material designers and researchers. Supporters of text simplification maintain that the process of simplification will increase the reader's ability to understand and interact with a text (Goodman & Freeman, 1993). Detractors argue that the removal of linguistic forms in favor of more simplified and frequent forms must inevitably deny learners the opportunity to learn the

natural forms of language (Long & Ross, 1993). Since the arguments from both sides have important connections to our current study, we will briefly describe them below.

a Supporters: Many researchers in second language acquisition have come to believe that the cognitive mechanisms involved in language acquisition are based on orientations similar to those located in simplified texts (e.g. increased frequency and reduced complexity; Tweissi, 1998). These cognitive mechanisms mimic the language found in simplified spoken input and help the language learner acquire a language in a relatively structured manner while avoiding unnecessary and distracting idiosyncratic style and maintaining communication features (Allen & Widdowson, 1979). Simplified texts have also been found to contain increased redundancy (Crossley et al., 2007; Crossley & McNamara, 2008), which is important in developing a cohesive text. Additionally, simplified texts are often seen as valuable aides to learning because they accurately reflect what the reader already knows about language and allow learners to extend this knowledge (Davies & Widdowson, 1974).

b Detractors: Many researchers argue that the process of simplification might actually complicate messages more than simplify them as a result of splitting complex sentences into independent sentences (Davison & Kantor, 1982). Such simplification processes might render the text more difficult to understand because shorter sentences may not allow the reader to engage in processes that assist comprehension, such as maintaining clear references to unfamiliar concepts, deleting irrelevant details in distracting phrases, removing pronouns with unclear antecedents, or highlighting important topics (Long & Ross, 1993). Researchers also argue that the dependence on simpler and more common words in English might correspond to the use of highly polysemous words, which are more ambiguous (Davies & Widdowson, 1974), thus leading to texts that are more difficult to process. Lastly, texts that are developed using traditional readability formulas may omit connectives between sentences in order to shorten the text's length to make it more readable. Such attempts often result in a text that is [more](#) difficult to comprehend (Goodman & Freeman, 1993).

In summary, then, the effects of text simplification produce texts that vary in terms of text cohesion and text sophistication. Supporters of text simplification argue that simplified texts provide more comprehensible input because they contain less sophisticated lexical features (more frequent words and simplified syntax) and increased cohesion through redundancy. Detractors argue that the process of simplification can lead to more complex linguistic features (more ambiguous words and shorter sentences that do not allow for proper referencing) and texts with less cohesion resulting from the omission of connectives.

2 The effects of simplification

While theories advancing the benefits or disadvantages of simplified texts are important, actual evidence for the strength of simplified texts lie in experimental studies. For instance, Long and Ross (1993) compared comprehension scores of L2 readers using three versions of the same text: authentic, simplified, and elaborated versions. Long and

Ross found that students who read the linguistically simplified text scored significantly higher on multiple choice items meant to test comprehension than did those that read the authentic version. They also found that students who read the elaborated version did not score significantly better than those that read the authentic or the simplified text. These results were replicated in a follow-up study (Yano et al., 1994), which demonstrated that there were no significant differences between simplified and elaborated texts, but simplified texts showed significant gains in reading comprehension over authentic texts, whereas elaborated texts did show gain over authentic texts. A later study conducted by Tweissi (1998) also found that simplification positively affected student's reading comprehension. However, Tweissi found that it was the type of simplification and not the amount of simplification that played the most important role in assisting the students to understand the material (e.g. texts with simplified lexicon led to greater comprehension gains than did other types of text modifications). Overall, these studies support the basic notion behind simplified texts: that the use of simplified input results in more comprehensible language.

3 Computational analyses of text differences

Other approaches to examining the effects of text simplification that have recently demonstrated success involve the use of corpus and computational analyses. Such approaches have proven especially profitable in examining the differences between simplified and authentic texts (Crossley et al., 2007; Crossley & McNamara, 2008). Crossley et al. used the computational tool Coh-Metrix to examine a wide range of linguistic features in authentic and simplified texts at the beginning and intermediate level. Crossley and McNamara (2008) found that at the intermediate level, authentic texts were syntactically more complex and contained a greater density of logical connectors. Simplified texts, on the other hand, had significantly higher levels of cohesion (e.g. lexical co-reference and semantic overlap) and lower levels of lexical sophistication (e.g. word frequency, word hypernymy, word polysemy and word familiarity). These findings were generally similar to those at the beginning level (with the exception of syntactic complexity; Crossley et al., 2007) and suggest that simplified texts may be more cohesive than authentic texts and contain less sophisticated linguistic features. These findings thus support the notion that the process of simplification creates texts that contain more linguistic features related to text comprehension.

One important limitation of these studies is that the investigated corpora were not directly comparable across levels. In other words, the texts in both the simplified and authentic corpora were from different sources. Since the texts used in the Crossley et al. studies came from general grammar and reading textbooks, it is also unclear which simplification approach authors used to guide text modification (structural or intuitive). Such a distinction is necessary if we are to understand the linguistic effects these approaches have on L2 reading texts. Although these corpus and computational studies have improved our understanding of the linguistic characteristics of simplified texts as compared to authentic texts, there is still considerable work to be done before we can claim to understand how the process of simplification alters the linguistic features of text.

Table 1. Descriptive statistics for the simplified corpora

	Mean number of words	Standard deviation	Mean number of paragraphs	Standard deviation
Beginner	636.450	162.239	9.820	2.935
Intermediate	703.140	164.082	10.110	2.934
Advanced	765.790	165.124	10.500	2.934

II Method

In order to examine the linguistic differences among levels of simplification, we collected a corpus of intuitively simplified reading texts designed for L2 instruction (Allen, 2009). We analysed the texts using the computational tool Coh-Metrix and used the results from the tool as a foundation for a statistical analysis of the reading texts. Since our interest is in determining the effects of intuitive text simplification, we selected texts used in L2 reading tasks that had been intuitively simplified at three different levels: beginner, intermediate, advanced. Using corpus, computational, and statistical approaches, this study examines which linguistic features distinguish the levels of the L2 reading texts from one another. We use the findings from this analysis to examine the linguistic structure of the simplified texts and consider the implications the findings have for the role of comprehensible input (e.g. text cohesion and text sophistication) and intuitive simplification methods.

I Corpus selection

The corpora used for this study is an extended version of the corpus used previously in Allen (2009). The texts that make up the corpus are 300 simplified L2 reading texts taken from an English teaching website (Onestopenglish, 2007). The texts are simplified versions of 100 authentic texts. The website provides simplified news texts and accompanying learning activities. The news texts are originally taken from the *Guardian Weekly*, a British-based publication with a wide international readership. The articles are originally selected by the website editors for their topicality and interest value and typically center on world affairs. These news texts are then simplified to three levels of simplification (beginning, intermediate, and advanced) by a team of authors. The method of simplification is intuitive, that is, without recourse to word lists or structural grading schemes (for more details, readers are referred to Allen, 2009). Since we are interested in distinguishing levels of simplification from among one another and not from authentic texts (from which we assume the modified text differ), our corpus only contains the simplified reading texts. Thus, our corpus contains 100 texts at each level of analysis. The total size of our simplified news corpus is 210,538 words. The total sizes of the sub-corpora are as follows: advanced = 76,579 words; intermediate = 70,314 words; beginning = 63,645. The sizes of the sub-corpora reflect the differences in text length due to the abridging of text at lower levels. Descriptive statistics for the corpus are located in Table 1.

2 Statistical analysis

To examine the hypothesis that there are a variety of different linguistic features that differentiate levels of simplified texts, we conducted a discriminant function analyses (DFA). A DFA is a common approach used in many previous studies that distinguish text types (e.g. Biber, 1993; Crossley & McNamara, 2009). The DFA in this study analysed the different levels of simplified texts in the corpus (the beginning, intermediate and advanced L2 reading texts) to examine what linguistic features distinguish the texts from one another.

For the statistical analysis, we selected Coh-Metrix indices that measured linguistic features related to cohesion (e.g. connectives, word overlap, semantic co-referentiality), linguistic sophistication (e.g. lexical difficulty and syntactic complexity) and surface level variables (e.g. number of words in the text). The indices selected for this study come from measures of lexical co-referentiality, semantic co-referentiality (Latent Semantic Analysis), word frequency, lexical diversity, word information taken from the MRC Psycholinguistic Database, hypernymy and polysemy values taken from WordNet, spatial cohesion, and causal cohesion, temporal cohesion and syntactic complexity. Many of these indices measure only explicit linguistic features and do not capture the entirety of the dimension to which they are related. The text comprehensibility and readability constructs connected to these measures are briefly discussed below. Detailed information about how the measures are calculated can be found in Graesser et al. (2004) and Crossley and McNamara (2009). The majority of the selected indices have been used in past studies to distinguish between simplified and authentic L2 reading texts (Crossley et al., 2008; Crossley & McNamara, 2009). All Coh-Metrix variables are normalized for text length; thus, the text length differences noted in corpus selection are not reflected in the Coh-Metrix analysis.

a Lexical co-referentiality: Coh-Metrix measures lexical co-reference between sentences by calculating noun overlap between sentences, argument overlap between sentences, stem overlap between sentences, and content word overlap between sentences. Lexical co-reference is an important indicator of cohesion that has been shown to aid in text comprehension and reading speed (Rashotte & Torgesen, 1985). Word overlap is also an important indicator of a text's readability (Crossley et al., 2008).

b Latent Semantic Analysis (LSA): Coh-Metrix tracks semantic co-referentiality using LSA, which is a mathematical and statistical technique for representing deeper world knowledge based on large corpora of texts. Semantic co-referentiality is an important indicator of text cohesion (Landauer et al., 2007). Coh-Metrix also tracks given/new information using LSA. Given information is less taxing on the reader's cognitive load (Chafe, 1975) and thus promotes greater text comprehension.

c Word frequency: The primary word frequency count in Coh-Metrix comes from CELEX (Baayen et al., 1995), the database from the Centre for Lexical Information, which consists of word frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9-million-word corpus. Measuring word frequency is important because

more frequent words allow for quicker decoding (Rayner & Pollatsek, 1994). Frequent words are also processed more quickly and more likely to be understood than infrequent words (Haberlandt & Graesser, 1985). The decoding and processing effects of frequent words also plays an important role in L2 text readability (Crossley et al., 2008).

d Word information (MRC Psycholinguistic Database): Coh-Metrix calculates information at the lexical level on five psycholinguistic matrices: familiarity, concreteness, imagability, meaningfulness and age of acquisition. All of these measures come from the MRC Psycholinguistic Database (Coltheart, 1981). Many of these indices are important indicators of lexical sophistication (Crossley & McNamara, 2009; Salsbury et al., 2011). The MRC word meaningfulness score relates to how strongly words associate with other words, and how likely words are to prime or activate other words. MRC word familiarity, concreteness, imagability and age of acquisition scores measure lexical constructs such as word exposure (familiarity), word abstractness (concreteness), the evocation of mental and sensory images (imagability), and intuited order of lexical acquisition (age of acquisition).

e WordNet indices: Coh-Metrix uses the WordNet database (Fellbaum, 1998) to calculate hypernymy and polysemy values. Coh-Metrix measures the relative ambiguity of a text by calculating its average lexical polysemy value (i.e. the number of meanings attributable to each word in text). Coh-Metrix measures the relative specificity of the words in a text by calculating its average lexical hypernymy value (i.e. the level of each word in a text within a conceptual, taxonomic hierarchy).

f Spatiality: Coh-Metrix represents motion spatiality through motion verbs and location spatiality through location nouns and location prepositions (Dufty et al., 2006). Coh-Metrix retrieves classifications for both motion verbs and location nouns from WordNet (Fellbaum, 1998). Coh-Metrix uses this information to produce a variety of indices related to spatiality. Key among these is the ratio of location and motion words, the number of locational nouns, the number of locational prepositions, the number of motion verbs and number of locational prepositions. Spatial cohesion helps to construct a text and ensures that the situational model of the text is well structured and clearly conveys text meaning (Zwaan et al., 1995).

g Temporal cohesion: Temporal cues also help construct a more coherent situation model of a text (Zwaan et al., 1995). There are three principal measures in Coh-Metrix related to temporality: aspect repetition, tense repetition, and the combination of aspect and tense repetition. Tense helps to organize events along timelines and can affect the activation of information in working memory. Tense also relates lexical events to a certain point in time. Aspect, on the other hand, conveys the dynamics of the point itself such as the point being ongoing or completed (Klein, 1994). Aspect helps maintain information in working memory (Magliano & Schleich, 2000).

h Causal cohesion: Along with spatiality and temporality, causality helps to construct the situation model of a text (i.e. the mental representation of a text; Zwaan et al., 1995). 'Causal cohesion' is measured in Coh-Metrix by calculating the ratio of causal verbs to

causal particles, which relates to the conveyance of causal content. The causal verb count is based on the number of main causal verbs identified through WordNet (Fellbaum, 1998). Causality is relevant to texts that depend on causal relations between events and actions (i.e. stories with an action plot or science texts with causal mechanisms) and is also relevant at the sentential level in order to convey causal relationships between simple clauses (Pearson, 1974–75).

i Lexical diversity: Lexical diversity is generally measured using type–token ratios (TTR). TTR is the division of types (i.e. unique words occurring in the text) by tokens (i.e. all instances of words). However, TTR indices are generally highly correlated to text length and are not reliable across a corpus of texts where the token counts differ markedly (McCarthy & Jarvis, 2007). To correct for the problem of text length in LD indices, a wide range of more sophisticated approaches to measuring vocabulary range have been developed. Those reported by Coh-Metrix include *MTLD* (measure of textual, lexical diversity; McCarthy, 2005) and *D* (Malvern et al., 2004) values.

j Syntactic complexity: Coh-Metrix assesses syntactic complexity in three ways:

- calculating the mean number of words before the main verb;
- measuring the mean number of high level constituents (sentences and embedded sentence constituents) per word; and
- measuring the uniformity and consistency of the syntactic constructions in the text.

More uniform syntactic constructions result in less complex syntax that is easier for the reader to process. Sentences that have less syntactic complexity are easier to process and comprehend (Perfetti et al., 2005). Sentences that contain more syntactic similarity lead to more readable texts for L2 learners (Crossley et al., 2008).

k Number of words in the text: We include a measure of text length because text length can correlate to the number of propositions contained in a text. The number of propositions found in text relates to text comprehension because of the strain multiple propositions put on working memory (Kintsch & Keenan, 1973).

3 Variable selection

To select the variables from the chosen Coh-Metrix indices, we used an analysis of variance (ANOVA) to examine which variables demonstrated the greatest differences between the levels of the reading texts. The selected variables were later used in a DFA to predict the text reading level.

III Analysis and results

We are interested in investigating what linguistic features distinguish among simplified reading text levels. That is to say, we want to examine if the reading levels in our corpus exhibit linguistic differences among each other. Such an approach will allow us to

examine the process of intuitive text simplification and to evaluate the effectiveness of the simplification process in terms of text comprehensibility and readability.

1 ANOVA

An ANOVA was conducted using the selected Coh-Metrix measures as the dependent variables and the essays from the training set as the independent variables. We selected the variable with the largest effect size as the representative variable for that measure. Considering that the training set contained 300 writing samples and using a conservative estimate of one predictor per 15 to 20 items, we determined that 15 indices would be an appropriate number of predictors for the discriminant analysis that would not create problems of overfitting. Such a ratio is standard for analyses of this kind (Field, 2005). We selected more than one variable from the MRC Database and WordNet because the indices measured different linguistic features. Descriptive statistics for the selected variables are presented in Table 2. All measures except temporal cohesion and connectives reported at least one index that demonstrated significant differences between the L2 reading text levels.

2 Pearson correlations

To avoid the risk of colinearity between variables, which would waste potential model power, we tested the variables for colinearity to ensure that no index pair correlated above $r \geq .70$. If variables that correlated above .70 were used in the model, it would make interpretation difficult because it would be unclear which variables were contributing to the model because many of the variables might be redundant (Brace et al., 2006). Pearson correlations showed that the Word Imagability and Word Concreteness indices were highly correlated with the Word Meaningfulness index ($r = .806, p < .001$ and $r = .718, p < .001$ respectively). The Word Meaningfulness index was retained in the analysis because it reported higher effect sizes than did the other two indices. Pearson correlations also revealed that the LSA Given/New index was highly correlated with the LSA sentence to sentence index ($r = .901, p < .001$). Because the LSA Given/New index reported higher effect sizes, it was retained in the analysis while the LSA sentence-to-sentence index was dropped.

3 Pairwise comparisons

As part of the ANOVA, a series of pairwise comparisons were conducted to examine differences between the L2 reading texts. Those results are reported below.

a Lexical diversity: The findings show that all levels of the L2 reading texts showed significant differences in lexical diversity values. Advanced reading texts had the highest lexical diversity values, while beginning level texts had the lowest.

b CELEX word frequency: The findings demonstrate that all levels of the L2 reading texts showed significant differences in word frequency values. Advanced reading texts had the lowest word frequency values, while beginning level texts had the highest.

Table 2. Means (standard deviations), F values, and effect sizes for simplified texts

Variables	Beginner	Intermediate	Advanced	<i>F</i> (2,297)	<i>hp</i> ²
Lexical diversity D	120.920 (29.297)	155.210 (30.514)	184.760 (34.838)	101.967	0.407
CELEX word frequency content words	2.423 (.131)	2.318 (.128)	2.213 (.120)	68.823	0.317
Word familiarity content words	576.198 (7.414)	571.708 (7.271)	567.258 (7.130)	37.780	0.203
Word meaningfulness every word	356.436 (11.104)	351.032 (11.157)	347.091 (11.178)	17.714	0.107
Total number of words	636.450 (162.238)	703.140 (156.311)	765.790 (183.387)	14.873	0.091
Number of words before the main verb	4.417 (1.037)	4.995 (1.271)	5.297 (1.356)	13.229	0.082
Verb hypernymy	1.444 (.016)	1.483 (.134)	1.542 (.132)	12.221	0.076
LSA given/new	0.314 (.040)	0.298 (.044)	0.285 (.044)	11.275	0.071
Number of motional prepositions	39.327 (11.776)	43.220 (9.757)	46.022 (9.954)	10.191	0.064
Word imagability every word	340.400 (11.092)	335.788 (10.960)	333.711 (10.544)	9.924	0.063
Sentence syntax similarity	0.101 (.019)	0.090 (.026)	0.086 (.028)	9.869	0.062
Word concreteness every word	314.445 (11.197)	310.040 (10.851)	308.394 (10.026)	8.546	0.054
LSA sentence to sentence	0.165 (0.066)	0.150 (.065)	0.133 (0.588)	6.439	0.042
Noun overlap	0.320 (.144)	0.286 (.129)	0.255 (.121)	6.170	0.040
Incidence of negation connectives	9.714 (5.497)	8.561 (5.033)	7.360 (4.187)	5.684	0.037
Ratio of causal verbs and particles	34.261 (8.052)	32.280 (8.691)	30.769 (9.727)	3.914	0.026
Word polysemy	3.733 (.325)	3.721 (.290)	3.625 (.287)	3.905	0.026

c Word familiarity: The results from the word familiarity scores demonstrate that all levels of reading texts show significant differences. The scores reveal that the beginning level texts contain more familiar words, while advanced texts contain less familiar words.

d Word meaningfulness: The findings show that all levels of text demonstrate significant differences in the use of meaningful words. The advanced texts had the highest word meaningfulness values, while beginning texts had the lowest.

e *Number of words*: The findings from the number of words index demonstrate that all text levels differed in their number of words. As expected, the beginning level texts had the fewest words and the advanced texts had the most words.

f *Syntactic complexity*: The findings from the number of words before the main verb index demonstrate that all text level comparisons except intermediate to advanced texts showed significant differences in syntactic complexity. Beginning texts had the fewest words before the main verb and advanced texts had the most.

g *Verb hypernymy*: The findings from the verb hypernymy index demonstrate that all text levels showed significant differences. Beginning level texts contained the least specific verbs while advanced texts contained the most specific verbs.

h *LSA givenness*: The findings from the given/new index show that all text levels demonstrated significant differences. Beginning level essays contained the most given information while advanced level texts contained the most new information.

i *Motional prepositions*: The findings from the motional prepositions index demonstrate that all levels of reading texts showed significant differences. Beginning level texts contained the fewest motional prepositions, while advanced texts contained the most.

j *Sentence syntax similarity*: The findings from sentence syntax similarity demonstrated that all levels of simplification showed significant differences between each other. Beginning level texts had the most syntactic similarity and advanced texts had the least.

k *Noun overlap*: The results of the noun overlap values demonstrate that significant differences exist between beginning and advanced level texts, but not with intermediate level texts. Beginning level texts contained the most noun overlap, while advanced texts contained the least.

l *Logical operators*: The results demonstrate that all text levels except intermediate to advanced texts significantly differ in their use of negative operators (e.g. *not*, *cannot*, *no*, *neither*). The beginning level texts contained the most negation operators, while the advanced texts contained the fewest.

m *Causal cohesion*: The results from the causal cohesion analysis show significant differences between beginning and advanced level texts, but no significant differences with intermediate texts. Advanced level texts exhibited less causal cohesion than beginning level texts.

n *Polysemy*: The polysemy results demonstrate significant differences between beginning and advanced level texts, but not intermediate texts. Beginning level texts contained the most ambiguous words, while advanced texts had the least.

Table 3. Predicted text type vs. actual text type results from total and cross-validated set

Actual text type	Predicted text type		
	Beginner	Intermediate	Advanced
<i>Total set:</i>			
Beginner	80	17	3
Intermediate	18	61	21
Advanced	2	21	77
<i>Cross-validated set:</i>			
Beginner	78	18	4
Intermediate	22	49	29
Advanced	3	25	72

4 Accuracy of model

To test the accuracy of these indices to distinguish between the levels of L2 reading texts, we conducted a discriminant function analysis. A discriminant analysis is a statistical procedure that is able to predict group membership (the level of the reading text) using a series of independent variables (the selected Coh-Metrix variables). The DFA generates a discriminant function. The discriminant function acts as the algorithm that predicts group membership. This is done first with the entire set. Later, the discriminant function analysis model from the entire set is used to predict group membership of the essays using repeated cross-validation. In cross-validation a fixed number of folds, or partitions of the data, is selected. Once the number of folds has been selected, each fold is used for testing the model. We selected a one-fold cross-validation model in which one instance in turn is left out and the remaining instances are used as the training set (in this case the 299 remaining texts). The accuracy of the model is tested on the model's ability to predict the omitted instance. This allows us to test the accuracy of the model on an independent data set. If the results of the discriminant analysis in both the entire set and the n -fold cross-validation set are statistically significant, then the findings support the predictions of the analysis (that linguistic differences can be used to distinguish between reading text levels). We report the findings using an estimation of the accuracy of the analysis. This estimation is made by plotting the correspondence between the actual texts and the predictions made by the discriminant analysis (see Table 3).

The results demonstrate that the discriminant analysis correctly allocated 218 of the 300 essays in the training set ($df = 4, n = 300) \times 2 = 226.194, p < .001$) for an accuracy of 72.7% (*weighted Kappa value* = .674). For the cross-validated set, the discriminant analysis correctly allocated 199 of the 300 essays for an accuracy of 66.3%. Chance for both of these analyses is 33.3%.

This study also reports its results in terms of recall and precision. Recall scores are computed by tallying the number of hits over the number of hits plus misses. Precision is the number of correct predictions divided by the number of incorrect predictions. The differences between recall and precision are important because an algorithm could predict everything to be a member of a single group and score 100% in terms of recall.

Table 4. Precision and recall finding

Text set	Precision	Recall	F1
<i>Total set:</i>			
Beginner	0.800	0.800	0.800
Intermediate	0.616	0.610	0.613
Advanced	0.762	0.770	0.766
<i>Cross-validated set:</i>			
Beginner	0.757	0.780	0.769
Intermediate	0.533	0.490	0.511
Advanced	0.686	0.720	0.703

However, this could only happen by claiming members of the other group. If this were the case, then the algorithm would score low in terms of precision. By reporting both values, we can better understand the accuracy of the model. The accuracy of the model for predicting the level of the simplified reading texts can be found in Table 4. The accuracy of the model for the total set was .726. The accuracy for the cross-validated set was .660. The results provide strong evidence that the linguistic features can be used to classify levels of simplified reading texts.

IV Discussion

This study has demonstrated that significant differences in linguistic features related to text sophistication and text cohesion result across levels as a result of intuitive simplification processes. The majority of these differences support the notion that beginning simplified texts contain more text features related to comprehensible input than advanced simplified texts in that beginning simplified texts contain more cohesive devices and less sophisticated language than advanced simplified texts. Thus, we can state, with some confidence, that intuitive text simplification processes produce texts that become linguistically more comprehensible as the text level decreases. However, in a few cases, intuitive text simplification may produce linguistic features that render a text less comprehensible (e.g. spatiality and word polysemy); although there are relatively few such instances and those reported tend to be associated with increased lexical sophistication and not text cohesion.

The greatest effect sizes for differences between levels were generally found in the linguistic indices related to text sophistication. The majority of the indices that demonstrated strong relationships among the reading levels were lexical in nature. Of these, lexical diversity showed a large effect size ($hp2 > .35$; Cohen, 1992), while two other indices – word frequency and word familiarity – showed medium effect sizes ($hp2 > .15 < .35$). Other lexical indices that demonstrated significant differences among the text levels included word meaningfulness, verb hypernymy, word imaginability, word concreteness and word polysemy. In general, these indices indicate that as the level of text simplification in our corpora increases so does the level of lexical sophistication. For instance, at the beginning level, simplified texts contain less lexical variation and fewer

infrequent words than advanced texts. At the same time, beginning level simplified texts contain more familiar, meaningful and imageable words than advanced texts. All of these linguistic modifications should allow lower level simplified texts to be decoded more quickly than advanced simplified texts. From a syntactic perspective, beginning simplified texts also contain fewer words before the main verb, likely making the texts syntactically simpler and easier to parse.

However, not all the linguistic modifications that occur as the result of text simplification lead to texts that contain more comprehensible input. Lexically, two modifications occur that might make the texts more difficult to process. The first is the predominance of verbs in beginning level texts that are less specific (i.e. lower hypernymy values or superordination) and thus potentially more abstract. L1 studies have demonstrated that superordinate words are more difficult to acquire, especially when compared to basic category words because they are more conceptually abstract (Rosch et al., 1976). L2 studies have also shown that superordinate words are produced later by L2 learners, partially because of their abstractness (Crossley et al., 2008). The hypernymy modification found in our simplified corpus, while unintentional, likely results from a dependence on frequent verbs that are less specific and have the potential for over-generalization (e.g. *be, go, have*). The second modification that might make beginning level simplified texts more difficult to process is the difference in polysemy values between beginning and advanced texts with beginning texts containing more ambiguous words (or words with more meanings). This modification, while also not intentional, also likely results from the use of more frequent words that have more meanings (Davies & Widdowson, 1974).

In reference to text cohesion, the findings of this study generally support the notion that as texts are simplified from advanced to beginning levels of readability, more cohesive devices are found in the texts. In reference to text similarity (or redundancy), the simplified texts demonstrated a progression of increasing similarity between sentence structures as the reading level decreased. For instance, beginning level texts had the greatest amount of syntactic similarity between sentences, meaning that the POS tags and the syntactic structures across sentences were more similar at the beginning level than at the intermediate level and more similar at the intermediate level than the advanced level. Beginning texts also contained more noun overlap and given information than advanced texts generating greater lexical redundancy. In reference to maintaining explicit links across texts, beginning texts demonstrated a greater use of negation connectives (*not, no, neither, nor*) and causal particles (*because, since, so, then*) and verbs than advanced texts. No significant differences, however, were found in the incidence of connectives. This was also true for indices of temporal cohesion.

The only potential loss of cohesion that occurs in our study was with spatial cohesion. Specifically, there were fewer motional prepositions in beginning level texts than in advanced level texts. This difference may make it more difficult for the reader to construct text meaning in relation to motional spatiality. However, motional prepositions are only one of many indices that help construct spatial cohesion. The other indices (motional verbs, locational nouns and locational prepositions) did not demonstrate significant differences. Thus, the overall effect on spatial cohesion that results from fewer motional prepositions is likely negligible.

The results of our statistical analysis demonstrated that our selected indices were best at classifying the texts from the beginning and advanced levels. However, while the reported Kappa value for the classification demonstrated substantial agreement between the DFA results and the actual text levels, a number of texts were misclassified by the model. For instance, three beginning level texts were classified as advanced, and two advanced texts were classified as elementary. Understanding why these texts were misclassified can help us better explore the process of intuitive simplification. For example, an analysis of the three beginning level texts that were classified as advanced demonstrates that two of the texts were much longer than the average length of the simplified texts in the beginning corpus (894 and 962 words respectively as compared to an average length of 637 words for beginning simplified texts). The additional length of these texts likely resulted in greater lexical coverage and fewer modifications (per word) than in shorter texts. This assumption is supported by lower scores for these texts in word meaningfulness, word familiarity, and word imageability. These differences along with the text length differences are the likely reasons the texts were misclassified. The two advanced texts that were classified as beginning level were both personal narratives (the first was autobiographical and the second was an interview). The nature of these texts likely influenced the linguistic features of the text. That is to say, the narrative style led to texts that were less formal in style and, as a result, contained more frequent words that were more familiar, meaningful and imageable. In the case of the personal interview, the lexical diversity value was much lower than average for advanced texts. This result is the product of the increased redundancy in the spoken speech found in the text.

The discriminant function analysis was least accurate at classifying the intermediate texts. The lower accuracy in intermediate texts is likely because these texts form a continuous transition from beginning to advanced levels, and are thus similar in features relative to the extreme levels. The notion that the intermediate level texts have similarities to both upper and lower levels (Allen, 2009) is the most plausible reason why the model used in this study was unable to accurately classify many of the intermediate texts. It is also the most plausible reason as to why there was sometimes little significant variation found between the intermediate level and other levels.

Overall, this study helps to support the strength of intuitive simplification processes. Considering that intuitive processes may be the most common approach to text simplification, it is important to understand their effects on the linguistic construction of the text and how they might affect text comprehensibility. The effects of intuitive text simplification also have implications for classroom teaching, second language acquisition and materials development. For instance, unlike text simplification processes that rely on mathematical algorithms (readability formulas), intuitive processes of simplification are, by definition, vague. Analyses that can illuminate these implicit processes benefit our understanding of how texts are processed and how this text processing influences materials design. Such analyses can also influence classroom teaching in that they better our understanding of text construction and how the features of a text affect the text's comprehensibility. Our analysis has demonstrated that intuitive text simplification leads to texts that are predominantly more cohesive and less sophisticated as the text level decreases and that the differences across levels follow expected trends. Such linguistic modifications should likely produce texts that are more comprehensible for beginning level

learners and allow those learners to interact with and comprehend the text to a greater degree. The modifications should also allow the texts to be decoded more quickly. The only unintended consequences that result from increasing text simplification appear to be the production of words that are potentially more ambiguous and abstract and the deletion of motional prepositions. All of these linguistic differences are all likely related to the dependence of more frequent words in more simplified texts. While these linguistic features may potentially disrupt the comprehensibility of the text, the advantages given by more frequent words should far outweigh their costs.

V Conclusions

This study has helped us to better define and understand the linguistic effects of intuitive text simplification. The study has also allowed us to see how the resulting effects may influence text comprehension and readability. In general, the simplified texts in the present corpus appear to provide suitable linguistic input for language learners at varying levels of proficiency, as demonstrated by reduced lexical sophistication and increased cohesion. These findings therefore provide support for the use of intuitively simplified texts when creating language learning materials. However, the nature of the intuitive approach entails that one group of authors' simplification procedures may differ considerably from another group. It is therefore important that authors be aware of the positive and negative implications of intuitive textual simplification, and that simplification procedures require attention and careful consideration.

There is also a need for additional corpus studies analysing a wider range of simplified texts and a wider range of authors. Such studies would allow for greater generalizability of the present findings. Finally, a study comparing texts simplified using structural approaches with those simplified using intuitive approaches would be helpful in examining potential differences between the approaches. Such a study would help us to further understand the processes of simplification, the resulting effects upon the linguistic features of texts, and the benefits or disadvantages of the respective approaches.

References

- Allen, D. (2009). A study of the role of relative clauses in the simplification of news texts for learners of English. *System*, 37, 585–99.
- Allen, J. & Widdowson, H.G. (1979). Teaching the communicative use of English. In: C. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 124–42). Oxford: Oxford University Press.
- Baayen, R.H., Piepenbrock, R., & van Rijn, H. (Eds.) (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium.
- Biber, D. (1993). Representativeness in corpus design. *Journal of Literary and Linguistic Computing*, 8, 243–57.
- Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologists: A guide to data analysis using SPSS for Windows*. 3rd edition. London: Palgrave.
- Carrell, P.L. (1987) Readability in ESL. *Reading in a Foreign Language*, 4, 21–40.
- Chafe, W.L. (1975). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: C.N. Li (Ed.), *Subject and topic* (pp 26–55). New York: Academic.

- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33A, 497–505.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–59.
- Crossley, S.A. & McNamara, D.S. (2008). Assessing L2 reading texts at the intermediate level: An approximate replication of Crossley, Louwrese, McCarthy & McNamara (2007). *Language Teaching*, 41, 409–29.
- Crossley, S.A. & McNamara, D.S. (2009). Computationally assessing lexical differences in second language writing. *Journal of Second Language Writing*, 17, 119–35.
- Crossley, S.A., Greenfield, J., & McNamara, D.S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42, 475–93.
- Crossley, S.A., Louwrese, M.M., McCarthy, P.M., & McNamara, D.S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91, 15–30.
- Davies, A. & Widdowson, H. (1974). Reading and writing. In: J.P. Allen & S.P. Corder (Eds.), *Techniques in applied linguistics* (pp. 155–201). Oxford: Oxford University Press.
- Davison, A. & Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17, 187–209.
- Dufty, D.F., Graesser, A.C., Lightman, E., Crossley, S.A., & McNamara, D.S. (2006). An algorithm for detecting spatial cohesion in text. Presentation at the 16th Annual Meeting of the Society for Text and Discourse, Minneapolis, MN, USA.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage.
- Gaies, S.J. (1983). The investigation of language classroom processes. *TESOL Quarterly*, 17, 205–17.
- Goodman, K. & Freeman, D. (1993). What's simple in simplified language. In: M.L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 69–76). Singapore: SEAMEO Regional Language Center.
- Graesser, A., McNamara, D., Louwrese, M., & Cai, Z. (2004). *Coh-Matrix: Analysis of text on cohesion and language*. *Behavioral Research Methods, Instruments, and Computers*, 36, 193–202.
- Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal*, 26, 5–24.
- Haberlandt, K.F. & Graesser, A.C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experiment Psychology: General*, 114, 357–74.
- Hatch, E. (1978). *Second language acquisition*. Rowley, MA: Newbury House.
- Hill, D. (1997). Survey review: Graded readers. *ELT Journal*, 51, 57–81.
- Hill, D. (2008). Graded readers in English. *ELT Journal*, 62, 184–204.
- Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257–74.
- Klein, W. (1994). *Time in language*. London: Routledge.
- Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.) (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- Little, D., Devitt, S., & Singleton, D. (1989). *Learning foreign languages from authentic texts: Theory and practice*. Dublin: Authentik.
- Long, M. & Ross, S. (1993). Modifications that preserve language and content. In: M.L. Tickoo (Ed.), *Simplification: Theory and application* (pp. 29–52). Singapore: SEAMEO Regional Language Center.
- Lotherington-Woloszyn, H. (1993). Do simplified texts simplify language comprehension for ESL learners? In: M.L. Tickoo (Ed.). *Simplification: Theory and application* (pp.140–54). Singapore: SEAMEO Regional Language Center.
- Magliano, J.P. & Schleich, M.C. (2000). Verb aspect and situation models. *Discourse Processes*, 29, 83–112.

- Malvern, D.D., Richards, B.J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills: Palgrave Macmillan.
- McCarthy, P.M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International*, 66, UMI No. 3199485.
- McCarthy, P.M. & Jarvis, S. (2007). *vocd*: A theoretical and empirical evaluation. *Language Testing*, 24, 459–88.
- Nunan, D. (1999). *Second language teaching and learning*. London: Heinle & Heinle.
- Onestopenglish (2007). News lessons [online]. Macmillan English Campus, London. Retrieved from: <http://www.onestopenglish.com> (November 2011).
- Pearson, P.D. (1974–75). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relationships. *Reading Research Quarterly*, 10, 155–92.
- Perfetti, C.A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In: M.J. Snowling & C. Hulme (Eds.). *The Science of Reading: A Handbook* (pp. 227–47). Oxford: Blackwell.
- Rashotte, C.A. & Torgesen, J.K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, 20, 180–88.
- Rayner, K. & Pollatsek, A. (1994). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Salsbury, T., Crossley, S.A., & McNamara, D.S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27, 343–60.
- Simensen, A.M. (1987). Adapted readers: How are they adapted? *Reading in a Foreign Language*, 4, 41–57.
- Tweissi, A.I. (1998). The effects of the amount and the type of simplification on foreign language reading comprehension. *Reading in a Foreign Language*, 11, 191–206.
- Yano, Y., Long, M., & Ross, S. (1994). Effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44, 189–219.
- Young, D.J. (1999). Linguistic simplification of second language reading material: Effective instructional practice? *The Modern Language Journal*, 83, 350–66.
- Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995). Dimensions of situation-model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386–97.