Identifying Linguistic Cues that Distinguish Text Types: A Comparison of First and Second Language Speakers

Scott A. Crossley, Max M. Louwerse and Danielle S. McNamara
(Mississippi State University and University of Memphis)

Crossley, Scott A, Max M. Louwerse and Danielle S. McNamara. (2008). Identifying Linguistic Cues that Distinguish Text Types: A Comparison of First and Second Language Speakers. ***Language Research*** **44.2, 361-381.**

The authors examine the degree to which first (L1) and second language (L2) speakers of English are able to distinguish between simplified or authentic reading texts from L2 instructional books and whether L1 and L2 speakers differ in their ability to process linguistic cues related to this distinction. These human judgments are also compared to computational judgments which are based on indices inspired by cognitive theories of reading processing. Results demonstrate that both L1 and L2 speakers of English are able to identify linguistic cues within both text types, but only L1 speakers are able to successfully distinguish between simplified and authentic texts. In addition, the performance of a computational tool was comparable to that of human performance. These findings have important implications for second language text processing and readability as well as implications for material development for second language instruction.

Keywords: second language reading, text readability, materials development, text processing, psycholinguistics, corpus linguistics, computational linguistics

1. Introduction

Recent research has demonstrated that text readability for second language (L2) learners is better measured using cognitively inspired indices based on theories of text comprehension as compared to readability formulas that depend on surface level linguistic features (Crossley, Greenfield & McNamara 2008). Such research is based on the notion that successful reading depends not on access to overt linguistic markers of difficulty such as letters per word and words per sentence, but on psycholinguistic processes. Psycholinguistic theories of reading tell us that reader processes such as decoding, syntactic parsing, and meaning construction assist the reader in developing mental models important for text comprehension and retention. However, little if any research has been conducted that analyzes the ability of L2 speakers to identify

linguistic cues in text. Thus, in this study, we examine L2 speakers' ability to identify linguistic cues within two types of reading passages which differ in linguistic features. We further examine whether these L2 speakers can distinguish between these two types of passages. As such we determine whether L2 speakers have implicit knowledge (i.e., ability to identify linguistic cues) or explicit knowledge (i.e., ability to distinguish between types of texts) of linguistic features in text. The results are expected to provide evidence concerning the differences between how L1 and L2 speakers process and comprehend text.

Our two types of passages used here are simplified and authentic passages within L2 instructional books intended for L2 English speakers. Authentic passages are those which are not edited from their original form (e.g., original excerpts from novels, newspaper articles, etc.). Simplified passages are those that have been manipulated with the express purpose of making them more digestible for L2 speakers. These simplifications include changes such as using more familiar words, more frequent words, more word overlap, more semantic co-referentiality, and less syntactic complexity. We use these two types of passages here because past research has demonstrated that the simplification process manipulates important linguistic cues related to reading processes (Crossley, Louwerse, McCarthy & McNamara 2007; Crossley & McNamara 2008).

Our first goal is to examine how L1 and L2 English speakers identify linguistic cues and differences between simplified and authentic texts. Our second goal is to compare these human evaluations of the two different text types to those of a computational tool. Our purpose is to demonstrate possible relationships between computational and human judgments. The investigation of how L2 speakers process textual information is important for understanding the linguistic competence of L2 speakers as well as the design and selection of L2 reading texts. Comparisons between human judgments of text types and those made computationally are expected to prove valuable for L2 reading research, and in particular, the development of theoretical frameworks related to reading comprehension and the support of cognitively inspired readability formulas.

1.1. Psycholinguistic Models of Reading

Reading processes involve mental actions on linguistic features (Graesser, Gernsbacher & Goldman 2003). Most psycholinguistic theories of reading comprehension divide the reading processing into decoding, syntactic parsing, and meaning construction (Perfetti 1985, Rayner & Pollatsek 1994). These multilevel representations help encode information and allow for meaningful comprehension to occur. They are argued to be the most viable framework for understanding L1 and L2 reading processing (Koda 2005). For instance, van

Dijk and Kintsch's (1983) model of discourse comprehension represents reading texts at three levels of language: lexical, syntactic, and textual. These levels are related to word identification, parsing, and referential mappings. When conflated, these levels help users form mental models of a text. These three levels and their importance in reading comprehension are discussed in detail below.

Decoding: Decoding relates to word identification (phonetically and semantically) and is a strong predictor of L2 reading performance (Day & Bamford 1998, Koda 2005). When decoding is automatic, it places lesser demands on a reader's working memory. If decoding is not automatic, then working memory processes are dedicated to decoding and not to comprehension. This can affect the retention of textual information because as working memory decays (Field 2004), opportunities for comprehension are lost. Thus, readers with better decoding skills are more accurate and faster readers. Decoding is also related to frequency effects because readers who are skilled at decoding are able to better identify low frequency words (Perfetti 1985, Rayner & Pollatsek 1994). Frequent words are also processed more quickly and understood better than infrequent words (Haberlandt & Graesser 1985, Just & Carpenter 1980).

Syntactic Parsing: Syntax is important for models of reading because decoding provides only a portion of the information necessary to understand text. How words are integrated into larger syntactic units such as phrases, clauses, and sentences must also be considered (Just & Carpenter 1987, Rayner & Pollatsek 1994). Syntactic parsing helps the reader link underlying relationships between concepts. These relationships serve as a temporary structure upon which to organize ideas into concepts (Just & Carpenter 1987). If the syntax of a sentence is complex, higher demands are placed on working memory processing, especially for less skilled readers (Perfetti et al. 2005). This is likely because less skilled readers cannot immediately construct the appropriate syntactic structures of a language (Rayner & Pollatsek 1994) and instead process texts word-by-word (Field 2004).

Meaning Construction/Referential Mappings: While decoding and parsing allow the reader to conceptualize the meaning of a sentence in working memory, meaning is also constructed in long term memory as the reader moves forward in a text. When links are made between concepts at the suprasentential level, larger units of meaning are constructed. Thus, the meaning of text does not reside solely in the sum meaning of the individual sentences, but in how those sentences are connected together (Just & Carpenter 1987, Rayner & Pollatsek 1994). Overlapping vocabulary and semantic co-referentiality have been found to be important aspects of reading processing

and can lead to gains in both text comprehension and reading speed (Douglas 1981, Kintsch & van Dijk 1978, McNamara & Kintch 1996, Rashotte & Torgesen 1985).

1.2. Readability Formulas

Traditional readability formulas, such as Flesch Reading Ease (1948) and Flesch-Kincaid (1975), measure word and sentence length in a text to produce a score that is an estimate of the text difficulty. Traditional readability formulas are widely accepted by the reading research community and they have much merit. They have, for instance, been relatively successful at judging L2 text readability (Greenfield 2004). However, they have also been widely criticized by reading researchers for not going beyond the surface form of the sentence. For example, traditional measures do not consider the idea units expressed in the text (Crossley, Greenfield & McNamara 2008; Kintsch, Welsch, Schmalhofer & Zimny 1990; McNamara, E. Kintsch, Butler-Songer & W. Kintsch, 1996), the cohesion of the text (Crossley et al 2008; Graesser, McNamara, Louwerse & Cai 2004; McNamara et al. 1996), nor its syntactic complexity, rhetorical organization, or propositional density (Brown 1998, Carrell 1987). Readability formulas have therefore been widely criticized by discourse analysts as being weak indicators of comprehensibility (Brown 1998, Carrell 1987, Davison & Kantor 1982). Unfortunately, the majority of the readability formulas used in the development of L2 reading texts depend on surface level text representations such as the number of words per sentence, syllables per word (Carrell 1987, Greenfield 2004, Schulz 1981), and vocabulary counts (Bamford 1984, Brown 1998, Gaies 1977). Few texts are assessed for deeper level linguistic features related to cognitive processing.

Discontent in L2 research with traditional readability formulas led to researchers eschewing the formulas and using human evaluations of text difficulty to determine text readability for simplified and authentic texts. Many L2 reading specialists (Bamford 1984; Hill 1997a, 1997b; Hill & Thomas 1988a, 1988b, 1989; Thomas & Hill 1993; Woolard 1988) resorted to conducting large-scale, extensive qualitative studies of simplified reading texts in an attempt to assess their readability. These assessments depended on the evaluative responses of reading specialists and L2 teachers in either placing texts within hierarchies of difficulty or assessing texts for specific linguistic factors (e.g., vocabulary, grammar, and density of information). While these studies were valuable, they tended to be based on the subjective assessment of the graders and the textual information provided by the publishers. Moreover, they were time consuming and constrained in the number of lexical, grammatical, and discourse features that they could assess; thus limiting the use of the measures for large corpora of L2 texts.

Recent research has begun to address the criticisms of researchers who question the use of traditional readability formulas for the development of L2 reading texts. For example, Crossley et al. (2008) demonstrated that a readability formula predicated on cognitively inspired indices of reading was superior in assessing text readability for L2 readers. Using three variables related to psycholinguistic reading processes (decoding, syntactic parsing, and meaning construction) taken from Coh-Metrix (Graesser et al. 2004), Crossley et al. demonstrated that features related to reading processes predicted text reading difficulty significantly better than traditional readability formulas. Importantly, their readability formula had both construct and content validity.

1.3. Computer Assessments of Text Difficulty

Recent advances in computational linguistics and corpus linguistics have made it possible to investigate, computationally, various measures of text and language features that supersede surface components of language and instead explore deeper, more global attributes of language (Graesser et al. 2004). One tool that has proven promising is *Coh-Metrix* (Graesser et al. 2004), which measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis using a variety of lexical databases, syntactic parsers, and semantic analyzers.

Coh-Metrix has not been previously compared to human judgments of text identification, but has proven to be successful in distinguishing linguistic features in simplified and authentic texts related to cohesion and reading processes. For example, results from a study by Crossley et al. (2007) demonstrated that beginning simplified texts used in L2 textbooks differ from authentic texts in that simplified texts had greater lexical and semantic coreferentiality (i.e., word and conceptual overlap), higher word frequency, and more syntactic complexity. Subsequent studies on intermediate L2 reading texts yielded similar results with deviations only in syntactic complexity, which was greater in authentic texts at the intermediate level (Crossley & McNamara 2008). In addition, Coh-Metrix indices related to lexical frequency, syntactic complexity, and lexical coreferentiality have been used to discriminate between simplified and authentic texts (Crossley, Louwerse, McCarthy & McNamara 2007). These results demonstrate how the use of computational tools can be of assistance to L2 reading researchers, material developers, and L2 publishers. Importantly, they also show how simplified and authentic texts differ in their lexical, syntactic, and meaning construction and how these differences might influence how the texts are processed and comprehended.

2. Current Study

This study consists of two related experiments. In Experiment 1, L1 and L2 speakers of English were asked to complete a survey that prompted them to evaluate selected texts as being either authentic or simplified based on linguistic criteria related to psycholinguistic theories of reading. This experiment investigates how well L1 speakers and L2 speakers of English are able to identify linguistic cues in simplified and authentic reading texts used in L2 instruction. It also investigates whether or not L1 and L2 speakers are able to accurately distinguish between the two types of texts. The second experiment compares the performance of L1 speakers to the computational tool Coh-Metrix. The purpose of the second experiment is to examine similarities between computational algorithms and human judgments. We suggest that if L2 speakers can identify linguistic indices related to mental models of reading, this would provide evidence that L2 speakers have access to similar processing mechanisms as L1 speakers. In turn, if computational judgments informed through cognitive concepts are similar to human judgments, this should prove valuable for the development of reading texts.

3. Experiment 1

3.1. Method

Participants: Thirty-two participants took part in this study. All thirty-two participants were English language teachers who had extensive instructional exposure to English. Twenty-one of the participants were females and eleven were males. Twenty-two of the participants were L1 speakers of English from the United States and ten were L2 speakers of English. Of the L2 speakers, eight were of Mexican descent. The remaining two participants were of European descent (Table 1). All L2 speakers of English rated themselves as either high-intermediate or advanced speakers of English.

Table 1. Participant demographic information

	Age	N
L1 Speakers	37.81	22
L2 Speakers	37.00	10
All Participants	37.56	32

Materials: We selected simplified and authentic texts used in beginning level L2 textbooks to populate the survey instrument used in this study. We

selected simplified and beginning texts because past research has demonstrated that such texts differ significantly in their use of linguistic features related to cognitive reading models (Crossley, Louwerse et al. 2007; Crossley et al. 2008). Thus, in examining readers' ability to discriminate between these texts, we assume that linguistic connections to cognitive reading processes can be fleshed out.

In L2 instruction, reading texts are simplified at the beginning and intermediate levels to make them more comprehensible for L2 learners and help prepare those learners for more advanced, authentic text (Young 1999). Proponents of simplified texts contend that beginning L2 learners benefit from texts that are lexically, syntactically, and rhetorically less dense than authentic texts. Many material writers and L2 specialists continue to emphasize the practical value of simplified texts, especially for beginning and intermediate L2 learners (Johnson 1981, 1982; Shook 1997). Simplifying texts is considered particularly important for shorter texts that are linguistically dense (Young 1999). Authentic texts, on the other hand, are any texts created to fulfill social purposes in the language community for which they were intended. Such texts include novels, newspapers, magazine articles, handbooks and manuals (Little, Devitt & Singleton 1989). Proponents for the use of authentic texts in the L2 classroom argue that these texts provide more natural language and naturally occurring cohesion than do simplified texts, which some claim engender unnatural discourse that increases reading difficulty (Crandall 1995).

For this study, we selected texts from the beginning level L2 texts used in Crossley, Louwerse et al. (2007), which demonstrated differences between authentic and simplified texts in reference to word frequency, syntactic complexity, and coreferentiality. Half of the texts were authentic and the other half were simplified. The simplified texts had a mean length of 150 words (SD = 26.95); the authentic texts had a mean length of 150 words as well (SD = 12.92). The texts, while not matched for topic, were matched based on size, level (all beginning), and genre. The genres selected included narrative, history, science, and personal letters. Ensuring that the texts were about the same length was important because multiple studies have affirmed that text length can influence reading comprehension (Leow 1993, 1997). The reading level is another important variable to control for because intermediate and advanced textbooks are more likely to contain authentic texts (Crossley, Louwerse et al. 2007) and simplified texts at the intermediate level often contain more complex linguistic features than at the beginning level (Crossley & McNamara 2008). Because the simplified texts, as well as the authentic texts, were taken from commercial textbooks, no information was available about what strategies were used in the simplification process. We assumed that the texts were simplified using the common practices of reducing information, simplifying vocabulary and structure, and controlling for discourse (Lotherington-

Wolosyzn 1993, Simensen 1987). Examples of simplified and authentic texts used in this study can be found in Appendix 1.

Procedure: A survey instrument first introduced participants to the distinctions between simplified and authentic texts and how the texts were used both instructionally and socially. These instructions were based on published guidelines for the creation of simplified texts and covered control of discourse, control of lexicon, and control of density and information (Lotherington-Wolosyzn 1993, Simensen 1987). Participants were not given specific examples of how these controls are used in actual text simplification, but were presented with examples of authentic and simplified texts to provide them with a general example of the differences.

After the introduction, participants read 10 passages, half authentic and half simplified texts, in random order with no more than one text per page. Two surveys were used in the study. Both surveys contained the same texts and evaluations, but the text order was reversed to control for ordering effects. After reading the text on the page, the participant answered 10 questions about the text. The first 2 questions assessed the reader's recall of the text. The third and fourth questions judged how authentic and simplified the text was using a 6-point Likert scale. The last 6 questions were presented in random order to control for ordering effects. All six questions used a 6-point Likert scale. Three of these questions evaluated the text based on linguistic features. These questions included judgments of grammar difficulty, word difficulty, and duplication of ideas and themes (meant to correlate with psycholinguistic features of reading: decoding, parsing, and meaning construction). The remaining 3 questions evaluated the consistency of the participant's answers by asking follow-up questions that measured the authenticity and the simplicity of the text in terms similar to questions three and four. These included 1) *How real did the text seem?*, 2) *Did the text seem manipulated?*, and 3) *Would you find the text outside of a second language textbook?* These questions were intended to highly correlate with questions about the authenticity and the simplification of the text and were used as a means of ensuring the validity of the survey. An example of the survey is located in Appendix 2.

3.2. Analyses

We first conducted Student *t*-tests using the text types as the grouping variable and the texts' readability scores as the dependent variable to examine differences in readability between the simplified and the authentic texts. We examined two readability formulas: Flesch Reading Ease Score, and the Coh-Metrix Readability Score. We predicted that simplified texts would demonstrate lower readability scores.

We analyzed the survey results for consistency by correlating specific questions that were meant to be highly interrelated. These included the follow-up questions about how realistic the text was (meant to correlate with authenticity), how manipulated the text seemed (meant to correlate with simplification), and the likelihood of finding the text outside of an ESL setting (meant to correlate with authenticity).

The survey results were compared to the actual text type being evaluated and the participant's responses were statistically evaluated using Analyses of Variance (ANOVAs) to assess whether the simplified texts were rated significantly differently from authentic texts. We conducted this analysis to evaluate participants' ability to accurately identify whether a text was simplified or not. We conducted similar tests on the ratings of authenticity, grammar difficulty, duplication of ideas and themes, and word difficulty. We subdivided these analyses by participants: L1 speakers, and L2 speakers.

3.3. Results

3.3.1. Text Readability

T-test results for Flesch Reading Ease approached significance t (1, 8) = -1.56, p = .08 with authentic texts having higher readability scores. *T*-test results for the Coh-Metrix Readability Formula also approached significance t (1, 8) = -1.65, p = .07 with authentic texts having higher reading scores (see Table 2).

Table 2. Descriptive statistics: readability scores.

		n	Mean	Std. Deviation
Flesch Reading Ease Score	Simplified	5	80.19	3.87
	Authentic	5	84.45	4.73
Coh-Metrix Readability Score	Simplified	5	21.89	7.21
	Authentic	5	28.40	5.12

3.3.2. Survey Validity

Authentic Texts: Findings demonstrated that participants showed consistency in their evaluation of authentic texts. Correlations between survey questions demonstrated expected positive and negative correlations. When participants evaluated authentic texts, their judgments of authenticity correlated negatively with their evaluation of a text's simplification (r = -.78, p<.001, N = 160) and manipulation (r = -.77, p<.001, N = 160). In contrast, the participants' judgments of authenticity correlated positively with how realistic the text seemed to be (r =.62, p<.001, N = 160) and the likelihood of finding the text outside of an ESL setting (r =.65, p<.001, N = 160). All other related correlations were also significant (Table 3). These findings help to reinforce the

validity of the survey design.

Table 3. Correlations for authentic texts all participants (all correlations significant at p < .001)

	Simplified	Realistic	Find Text Outside of ESL Book	Manipulation
Authentic	-0.78	0.62	0.65	-0.77
Simplified		-0.52	-0.59	0.69
Realistic			0.57	-0.62
Outside ESL				-0.64

Simplified Texts: Participants also showed consistency in their evaluation of simplified texts. Correlation between survey questions demonstrated expected positive and negative correlations. When participants evaluated a simplified text, their judgments of simplification correlated negatively with judgments of a text's authenticity ($r = -.77$, $p < .001$, $N = 160$), how realistic the text seemed to be ($r = -.44$, $p < .001$, $N = 160$) and the likelihood of finding the text outside of an ESL setting ($r = -.56$, $p < .001$, $N = 160$). In contrast, the participants' judgments of simplification correlated positively with manipulation ($r = .60$, $p < .001$, $N = 160$). All other related correlations were also significant (see Table 4). These findings also help to reinforce the validity of the survey design.

Table 4. Correlations for simplified text all participants (all correlations significant at p < .001)

	Simplified	Realistic	Find Text Outside of ESL Book	Manipulation
Authentic	-0.77	0.61	0.68	-0.70
Simplified		-0.44	-0.56	0.60
Realistic			0.61	-0.52
Outside ESL				-0.62

3.3.3. Text Evaluations

L1 Speakers: L1 speaker participants were able to accurately identify whether a text was more authentic, $F(1, 220) = 24.80$, $p < .001$, or more simplified, $F(1, 220) = 25.56$, $p < .001$. Authentic texts received higher scores of authenticity and lower scores of simplification, while simplified texts received higher scores of simplification and lower scores of authenticity. Additionally, L1 speaker participants were able to accurately identify authentic texts based on grammatical difficulty, $F(1, 220) = 30.71$, $p < .001$, with the grammar in authentic texts evaluated as more difficult and the grammar in simplified texts as less difficult. Additionally, L1 speakers were able to accurately identify simplified texts based on duplication of ideas and themes, $F(1, 220) = 21.60$, p

< .001, with simplified texts evaluated as having more duplication and authentic texts as less. Differences in word difficulty were significant as well, $F(1, 220) = 36.15$, $p < .001$, with the words in authentic texts evaluated as more difficult and simplified texts as less difficult (see Table 5).

L2 Speakers: L2 speaker participants were unable to identify whether a text was more authentic $F(1, 100) = 3.85$, $p = .053$ or more simplified, $F(1, 100) = 1.71$, $p = .20$, though patterns were in the expected directions and the findings for authentic texts approached significance. Authentic texts received higher scores of authenticity and lower scores of simplification and simplified texts received higher scores of simplification and lower scores of authenticity, but these differences were not significant. However, L2 speaker participants were able to accurately identify authentic texts based on grammatical difficulty, $F(1, 100) = 19.66$, $p < .001$, with the grammar in authentic texts evaluated as more difficult and the grammar in simplified texts as less difficult. Additionally, L2 speakers were able to accurately identify simplified texts based on duplication of ideas and themes, $F(1, 100) = 5.86$, $p < .05$, with simplified texts evaluated as having more duplication and authentic texts as less. Differences in word difficulty were significant as well, $F(1, 100) = -19.50$, $p < .001$, with the words in authentic texts evaluated as more difficult and simplified texts as less difficult (see Table 6).

Table 5. Means (standard deviations) and F values for simplified and authentic texts L1 speakers

Variables	Simplified Texts	Authentic Texts	*F(1,320)*
Judgment of Authenticity	3.09 (1.51)	4.20 (1.78)	24.80
Judgment of Simplicity	4.03 (1.42)	2.96 (1.69)	25.56*
Grammar Difficulty	2.11 (.85)	2.95 (1.31)	30.71*
Duplication of Ideas and Themes	3.92 (1.33)	3.06 (1.42)	21.60
Word Difficulty	2.19 (.88)	3.08 (1.27)	36.15*

Note: *Welch's *F*

Table 6. Means (standard deviations) and F values for simplified and authentic texts L2 speakers

Variables	Simplified Texts	Authentic Texts	*F(1,320)*
Judgment of Authenticity	3.46 (1.90)	4.20 (1.87)	3.85
Judgment of Simplicity	3.44 (1.90)	2.94 (1.93)	1.71
Grammar Difficulty	1.34 (.63)	2.50 (1.74)	19.66*
Duplication of Ideas and Themes	3.56 (1.91)	2.68 (1.71)	5.87
Word Difficulty	1.30 (.65)	2.48 (1.78)	19.50*

Note: *Welch's *F*

3.4. Discussion

This study demonstrates that both L1 and L2 speakers of English were able to identify individual linguistic variables related to models of text comprehension. However, the findings demonstrated that L2 speakers could not identify text types holistically as being either simplified or authentic.

The results of the first analysis are informative for a number of reasons. First, when given a text and asked to evaluate it holistically based on the authenticity or simplification of linguistic variables, the L1 speakers in this survey were able to discriminate between the two reading text types. L1 speakers were also able to judge that authentic texts exhibited more grammatical difficulty and word difficulty than simplified texts and that simplified texts exhibited more duplication of ideas and themes than authentic texts. Because these variables are correlated with models of reading comprehension and text discrimination, it can be argued that the ability to recognize these features and discriminate between texts that manipulate them allows for ease of text processing. Such a finding is expected in L1 speakers and helps to support the notion that reading models include linguistic aspects related to decoding, parsing, and meaning construction.

L2 speakers of English were not able to identify to discriminate between reading text types. However, L2 speakers were able to judge that authentic texts exhibited more grammatical difficulty and word difficulty than simplified texts and that simplified texts exhibited more duplication of ideas and themes than authentic texts. This demonstrates that while L2 speakers are likely aware of how individual linguistic structures differ, they appear unable to construct larger models of text types based on the individual linguistic features. Thus, unlike L1 speakers who seem capable of identifying individual linguistic features related to the reading process and conflating them into a more global model of text type, L2 speakers seem unable to make the transition between linguistic features used in constructing a reading model and the actual model. This might demonstrate that L2 speakers are unable to merge linguistic features into a psycholinguistic model of reading in a similar manner as L1 speakers.

L2 speakers also appear more likely to identify authentic texts as being authentic than simplified texts as being simplified. This could be the result of authentic texts containing more natural occurrences of overlap and frequency. This might make the identification of authentic texts easier as compared to simplified texts which manipulate these linguistic variables. If so, authentic texts are likely better placed to provide L2 speakers with linguistic clues that related to reading models than simplified texts. Thus, although simplified texts might provide for more comprehension (Johnson 1981, 1982; Shook 1997), authentic texts might allow for the development of more complete reading models. This study seems to indicate that L2 speakers can identify linguistic

variables related to reading models, but cannot combine them to identify text types that differ in readability. Manipulating texts seems to make this task more difficult.

4. Experiment 2

4.1. Method

The purpose of Experiment 2 is to compare L1 speakers' judgments to those obtained from computational indices. From Coh-Metrix, we selected only variables that represent linguistic features related to the psycholinguistic processes of reading and were part of Crossley et al.'s (2008) readability study. These included the linguistic indices of CELEX word frequency, syntactic similarity, and content word overlap. In addition to these indices' correlation with reading processes, past studies have also demonstrated that simplified and authentic texts differ significantly in their use of these variables in text construction (Crossley, Louwerse et al. 2007; Crossley, McCarthy & McNamara 2007; Crossley & McNamara 2008). These indices are briefly discussed below.

Celex Word Frequency: Coh-Metrix calculates word frequency information through CELEX frequency scores. The CELEX database (Baayen, Piepenbrock & Gulikers 1995) consists of frequencies taken from the early 1991 version of the COBUILD corpus, a 17.9 million-word corpus. For this study, the *CELEX frequency score for written words* was selected as the lexical level variable. The measure is related to lexical decoding.

Syntactic Similarity: The index *semantic similarity* available in Coh-Metrix measures the uniformity and consistency of parallel syntactic constructions in text. The index looks not only at syntactic similarity at the phrase level, but also at the part of speech level with the assumption that the more uniform the syntactic constructions are, the less complex the syntax will be to process. The measure is related to syntactic complexity.

Content Word Overlap: The index *content word overlap* measures how often content words overlap between two adjacent sentences. The measure does not include function words such as articles, conjunctions, and prepositions. The measure is related to meaning construction.

4.2. Results

We examined the correspondence between the ability of human subjects

and computational tools to evaluate text types. To do so, the L1 speakers' responses on the survey instrument were correlated with Coh-Metrix indices for the passages using Pearson Moment Product Correlations.

Word Difficulty: The correlation between participant answers and Coh-Metrix measurements of word difficulty based on CELEX word frequency was significant ($r = .779, p < .01, N = 10$).

Syntactic Complexity: A correlation between participant answers and Coh-Metrix measurements on grammar difficulty and sentence syntax similarity was not significant ($r = -.378, p > .05, N = 10$).

Coreferentiality: The correlation between participant answers for text duplication and Coh-Metrix measurements of content word overlap was significant ($r = -.619, p < .05, N = 10$).

4.3. Discussion

The results from the second study indicated that when textual differences were assessed based on linguistic features alone (syntactic similarity, word frequency, and coreferentiality), similarities between human and computer judgments were found. Specifically, Coh-Metrix word frequency and content word overlap findings significantly correlated to participants' evaluations of word difficulty and duplication of ideas and themes. However, when considering similarities between Coh-Metrix judgments of syntactic complexity and participants' evaluations of grammatical difficulty, no significant correlations were found. The first two findings provide promise that computational algorithms may evaluate linguistic features in a similar fashion to human judgments and might thus link to linguistic features common to models of reading processes and automatic text processing related to readability. These findings help provide additional construct validity for cognitive tools. Eventually, such tools should be able to replicate human judgments and provide a means for quicker and more accurate text evaluation.

There are several possible reasons why the correlations between the computational tool and human participants in reference to syntactic complexity were not significant. First, the computational algorithms used by Coh-Metrix, while founded on cognitive models of language processing, are still in their infancy and are not fully compatible to human linguistic skills. Second, in the case of this study, it is possible that the tool and the participants were measuring different indices of syntactic difficulty. It is feasible that while Coh-Metrix measures syntactic similarity as an index of syntactic complexity, human judgments at the surface level may associate syntactic difficulty with part of speech types or grammatical structure.

or grammatical structure.

5. Conclusions

This study has demonstrated that L1 and L2 speakers are able to identify individual features of text that are related to reading models, but that only L1 speakers can identify texts as being either simplified or authentic. Additionally, human judgments correlated with two of three Coh-Metrix measures providing evidence that computational tools can evaluate linguistic features in a manner similar to humans.

Admittedly, the number of participants as well as the number of texts used in this study were small and could be viewed as limiting the extendibility of the study. Furthermore, the number of texts used in this study limited the statistical analyses employed as the number of items examined was small. In addition, the use of professional language teachers as participants might be a limitation because studies have shown that language teachers may primarily attend to surface level features of texts instead of treating texts at the level of discourse (Zamel 1976). Lastly, the use of short texts (150 words) may limit the generalizability of the findings because they may restrict the use of metacognitive strategies used by the participants (Swaffar 1991, Kintsch & Van Dijk 1978) and affect the amount of data available for analysis. With these limitations in mind, future studies should consider a larger corpus of texts and a greater number of linguistic indices (this current study considers three indices out of the hundreds available through Coh-Metrix).

However, we argue that these limitations do not influence the overall conclusions that we draw from this study. As an exploratory analysis, the study demonstrates that humans are able to identify the linguistic features in texts related to reading processes and that a computational tool measures some of these features in a similar manner. These findings could prove to be important in understanding how L1 and L2 speakers assess text types using linguistic features. Additionally, the study indicates that L2 speakers differ from L1 speakers in their ability to differentiate between texts types. This could prove important in identifying differences in how L1 and L2 speakers process linguistic information and use that information to inform textual decisions. Lastly, the ability of a computational tool to evaluating the readability of texts in a manner similar to human judgment is important because it could allow for faster assessment of the quality and types of passages for L2 instruction. Future studies should develop the hypotheses presented in this paper by testing readability between L1 and L2 speakers in both simplified and authentic texts. Such a study could provide additional evidence for different reading and processing models between L1 and L2 speakers and examine the effects of such

models on reading comprehension.

References

Baayen, R. H., R. Piepenbrock and H. van Rijn, eds. (1993). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium.

Bamford, J. (1984). Extensive reading by means of graded readers. *Reading in a Foreign Language* 2, 218-260.

Brown, J. D. (1998). An EFL readability index. *JALT Journal* 29.2, 7-36.

Carrell, P. (1987). Readability in ESL. *Reading in a Foreign Language* 4, 21-40.

Crandall, J. (1995). The Why, What, and How of ESL Reading Instruction: Guidelines for writers of ESL reading textbooks. In P. Byrd, ed., *Material Writer s Guide*, 79-94. Heinle and Heinle Publishers.

Crossley, S. A., M. M. Louwerse, P. M McCarthy and D. S. McNamara. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal* 91.

Crossley, S. A., P. M. McCarthy and D. S. McNamara. (2007). Discriminating between second language learning text-types. In D. Wilson and G. Sutcliffe eds., *Proceedings of the 20th International Florida Artificial Intelligence Research Society*, 205-210. AAAI Press.

Crossley, S. A. and D. S. McNamara. (2008). Assessing second language reading texts at the intermediate level: An approximate replication of Crossley, Louwerse, McCarthy and McNamara (2007). *Language Teaching* 41.3, 409-429.

Crossley, S. A., J. Greenfield and D. S. McNamara. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly* 42.3, 475-493.

Davison, A. and R. Kantor. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly* 17, 187-209.

Day, R. R. and J. Bamford. (1998). *Extensive Reading in the Second Language Classroom.* Cambridge University Press.

Douglas, D. (1981). An exploratory study of bilingual reading proficiency. In Hudelson, S, ed., *Learning to Read in Different Languages*, 33-102. Washington Center for Applied Linguistics.

van Dijk, T. A. and W. Kintsch. (1983). *Strategies of Discourse Comprehension.* Academic.

Field, J. (2004). Psycholinguistics: The key concepts. New York, Routledge.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology* 32, 221-233.

Gaies, S. (1977). The nature of linguistic input in formal second language learning: Linguistic and communicative strategies. In H. Brown, C. Yorio and R. Crymes, eds., *TESOL* 77, 19-23. TESOL.

Graesser, A. C., M. A. Gernsbacher and S. R. Goldman. (2003). *The Handbook of Discourse Processes The Handbook of Discourse Processes.* Erlbaum

Graesser, A., D. S. McNamara, M. Louwerse and Z. Cai. (2004). Coh-Metrix: Analy-

sis of text on cohesion and language. *Behavioral Research Methods, Instruments and Computer* 36, 193-202.

Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal* 26, 5-24.

Haberlandt, K. F. and A. C. Graesser. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experiment Psychology: General* 114, 357-374.

Hill, D. (1997a). Graded (Basal) readers-Choosing the best. *The Language Teacher* 21, 21-26.

Hill, D. (1997b). Survey review: Graded readers. *ELT Journal* 51, 57-81.

Hill, D. and H. R. Thomas. (1988a). Survey review: Graded readers (Part 1). *ELT Journal* 42, 44-52.

Hill, D. and H. R. Thomas. (1988b). Survey review: Graded readers (Part 2). *ELT Journal* 42, 124-136.

Hill, D. and H. R. Thomas. (1989). Survey review: Seven series of graded readers. *ELT Journal* 43, 221-231.

Johnson, P. (1981). Effects of reading comprehension of language complexity and cultural background. *TESOL Quarterly* 15, 169-181.

Johnson, P. (1982). Effects on reading comprehension of building background knowledge. *TESOL Quarterly* 16, 503-516.

Jurafsky, D. and J. H. Martin. (2002). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.* Prentice-Hall.

Just, M. A. and P. A. Carpenter. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review* 87, 329-354.

Kincaid, J. P., R. P. Fishburne, R. L. Rogers and B. S. Chissom. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*, Research Branch Report 8-75, Naval Technical Training, U. S. Naval Air Station.

Kintsch, W. and T. A. van Dijk. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363-394.

Kintsch, W., D. Welsch, F. Schmalhofer and S. Zimny. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language* 29, 133-159.

Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition* 15, 333-355.

Leow, R. P. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learn in* 8, 151-182.

Little, D., Devitt, S. and D. Singleton. (1989). *Learning Foreign Languages from Authentic Texts: Theory and Practice.* Authentik.

Lotherington-Woloszyn, H. (1993). Do simplified texts simplify language comprehension for ESL learners? In M. L. Tickoo, ed., *Simplification: Theory and Appli- cation*, 140-154. SEAMEO Regional Language Center.

McNamara, D. S. and W. Kintsch. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes* 22, 247-288.

McNamara, D. S., E. Kintsch, N. Butler-Songer and W. Kintsch. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction* 14, 1-43.

Schulz, R. (1981). Literature and readability: Bridging the gap in foreign language reading. *The Modern Language Journal* 65, 43-53.

Shook, D. (1997). Identifying and overcoming possible mismatches in the beginning reader-literary text interaction. *Hispania* 80, 234-243.

Simensen, A.M. (1987). Adapted readers: How are they adapted? *Reading in a Foreign Language* 4, 41-57.

Swaffar, J. K. (1991). Language Learning is more than learning a language: Rethinking reading and writing tasks in textbooks for beginning language study. In B. Freed ed., *Foreign Language Acquisition Research and the Class*, 252-279. Heath.

Thomas, H. R. and D. Hill. (1993). Seventeen series of graded readers. *ELT Journal* 47, 250-267.

Woolard, S. (1988). The battering ram: Some observations on graded texts. *Modern English Teacher* 15, 19-20.

Young, D. J. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *The Modern Language Journal* 83, 350-366.

Zamel, V. (1976). Teaching composition in the ESL classroom: What we can learn from research on the teaching of English. *TESOL Quarterly* 10, 67-76.

Appendix 1

Example Simplified and Authentic Texts Used in Survey

Simplified Text	Authentic Text
People started to grow big pumpkins more than 100 years ago. About 20 years ago, Howard Dill decided to have a contest. He brought his biggest pumpkin to the contest. It weighed 438 pounds. Every year, people brought bigger pumpkins. Two years ago, another pumpkin grower set a world record. His pumpkin weighed 1,061 pounds! One of the pumpkin contests is at a festival called Pumpkinfest. At Pumpkinfest, people bring giant pumpkins and other giant vegetables. There is a contest for the biggest watermelon and the biggest squash. There is even a contest for the world's biggest flowers. Many of the giant fruits and vegetables weigh more than 500 pounds. Some of the flowers are more than 14 feet high! At Pumpkinfest, people walk around and look at the giant fruits and vegetables. They can even buy them after the contest and take them home. "There is only one problem," said a man who bought a giant pumpkin. "I can't get the pumpkin into my truck!"	You can probably think of some questions that the scientific method could help you answer. Perhaps you have a younger brother. You know that he is much harder to get along with when he misses his afternoon nap. But what is it about his behavior that makes him harder to get along with? How does he act differently? Let's use the steps of the scientific method to see how you could find the answer. Ask a question. "How does my little brother act differently when he misses his afternoon nap?" Gather information about the question. Watch your brother on days when he takes a nap and on days when he misses his nap. Ask your parents or brothers or sisters how he acts differently. Form a hypothesis. "My little brother has less patience in the evenings on days he misses his nap than on days he takes a nap." Test the hypothesis.

Appendix 2

People started to grow big pumpkins more than 100 years ago. About 20 years ago, Howard Dill decided to have a contest. He brought his biggest pumpkin to the contest. It weighed 438 pounds. Every year, people brought bigger pumpkins. Two years ago, another pumpkin grower set a world record. His pumpkin weighed 1,061 pounds!

One of the pumpkin contests is at a festival called Pumpkinfest. At Pumpkinfest, people bring giant pumpkins and other giant vegetables. There is a contest for the biggest watermelon and the biggest squash. There is even a contest for the world's biggest flowers. Many of the giant fruits and vegetables weigh more than 500 pounds. Some of the flowers are more than 14 feet high!

At Pumpkinfest, people walk around and look at the giant fruits and vegetables. They can even buy them after the contest and take them home. "There is only one problem," said a man who bought a giant pumpkin. "I can't get the pumpkin into my truck!"

1. What is the main theme of this text?

   _____

2. How much do normal pumpkins weigh?

   _____

3. Please rate the above text on Authenticity by circling a number.
   Least Authentic [1-------2-------3-------4-------5-------6] Most Authentic

4. Please rate the above text on Simplicity by circling a number.
   Least Simplified [1-------2-------3-------4-------5-------6] Most Simplified

5. How difficult is the grammar? Please circle a number.
   Not difficult [1-------2-------3-------4-------5-------6] Very difficult

6. Does the text seem to exhibit duplication of ideas and themes? Please circle a number.
   No duplication [1-------2-------3-------4-------5-------6] Much duplication

7. How realistic is the language in this text? Please circle a number.
   Not realistic [1-------2-------3-------4-------5-------6] Very realistic

8. How difficult do the words seem? Please circle a number.
   Not difficult [1-------2-------3-------4-------5-------6] Very difficult

9. Would you find this text outside of an ESL textbook? Please circle a number.
   Unlikely [1-------2-------3-------4-------5-------6] Very likely

10. Does the text seem to have been manipulated? Please circle a number.
    Not manipulated[1-------2-------3-------4-------5-------6] Very manipulated

Scott A. Crossley
Department of English
Mississippi State University
P.O. Box E.
Starkville, MS 39759, USA
E-mail: sc544@msstate.edu

Max M. Louwerse
Department of Psychology
University of Memphis
202 Psychology Building
Memphis, TN 38121, USA
E-mail: mlouwerse@memphis.edu

Danielle S. McNamara
Institute for Intelligent Systems
University of Memphis
Memphis, TN 38152, USA
E-mail: d.mcnamara@mail.psyc.memphis.edu

Received: August 24, 2008
Revised version received: November 18, 2008
Accepted: November 21, 2008