Discriminating between Second Language Learning Text-Types

Scott A. Crossley, Philip M. McCarthy*, and Danielle S. McNamara*

Department of English
Mississippi State University
Mississippi State, MS 39762

*Institute for Intelligent Systems
University of Memphis
Memphis. TN 38152
{scrossley, pmccarthy, d.mcnamara} @ mail.psyc.memphis.edu

Abstract

Text classification remains one of the major fields of research in natural language processing. This paper evaluates the use of the computational tool *Coh-Metrix* as a means to distinguish between seemingly similar text-types. Using a discriminant analysis on a corpus of second language reading texts, this paper demonstrates that *Coh-Metrix* is able to significantly distinguish authentic text-types from ones that have been specifically simplified for second language readers. This paper offers important findings for text classification research and for second language reading materials developers and second language teachers by demonstrating that moderate, shallow, textual changes can affect discourse structures.

Introduction

Computationally distinguishing groups of highly similar text-types serves a wide variety of linguistic fields. In text mining, distinguishing text-types facilitates the accuracy of data retrieval and text identification (Kao & Poteet, 2006; Louwerse et al., 2004). In natural language processing, identifying text-types facilitates parsers by predicting syntactical organization and lexical-semantic likelihood (Kessler, Nunberg, & Schutze, 1997). And in forensic linguistics, identifying text-types allows for identifying perpetrators of fraudulent or deceptive claims (Colwell, Hiscock, & Memon, 2002; Newman et al., 2001)

In this paper, we build on such research by demonstrating an approach that successfully distinguishes between two seemingly similar text-types: authentic second-language learning texts and simplified second-language learning texts. Such research serves all fields where the accurate identification of text-type or the better understanding of subtle textual differences is paramount. In this study, the research also specifically facilitates the field of second-learning reading by providing empirical evidence as to the type and extent of the differences between the two most prominent forms of reading material: authentic and simplified texts.

Second Language Reading Text-Types

In the field of second language reading, the simplification of text is common. Second language reading texts are simplified at the beginning and intermediate levels in order to make the text more comprehensible for second language learners and to help prepare those learners for more advanced, authentic texts (Young, 1999). Support for such simplification rests on second language acquisition theories and the linguistic nature of simplified texts. In general, researchers that support simplified material assume that such text benefits second language learners because it excludes unnecessary and distracting, idiosyncratic styles without suffering a loss of valuable communication features and concepts that are found in authentic text (Allen & Widdowson, 1979). These researchers also argue that simplified text can be a valuable aid to learning because it accurately reflects what the reader already knows about language (Davies & Widdowson, 1974) and contains increased redundancy and amplified explanation (Kuo, 1993). The simplification of second language reading text is so common that publishing houses and editorial staff provide writers with prescriptive guidelines regarding the linguistic construction of texts. These guidelines generally call for the control of information structure, the control of language, and the control of discourse (Simenson, 1987).

In opposition to such simplification is a movement toward the use of authentic text in the classroom. Authentic text is any text that was originally created to fulfill the social purpose in the language community for which it was intended. These texts include novels, newspapers and magazine articles, or handbooks and manuals (Little, Devitt, & Singleton, 1989). In supporting the use of authentic text, second language reading researchers often turn to the use of linguistic features and specifically to cohesive devices. Honeyfield (1977), for instance, suggests that modifications to authentic text affect the text's cohesion, making it simpler in appearance but more difficult for L2 readers to understand and manage. Researchers have also argued that the recognition and understanding of cohesive devices such as conjunctions and other inter-sentential linguistic devices by second language learners and readers is vital to the development of reading comprehension skills and information processing skills (Cowan, 1976; Mackay, 1979). With regard to the lexicon, researchers argue that good readers use the natural redundancy found in authentic text to their advantage by using the redundancy to help them reconstruct the entire text and understand unfamiliar

lexicon (Goodman, 1976; Johnson, 1982; Graesser et al., 2004).

In sum, proponents of authentic text in the second language classroom support their position by addressing the idea that authentic text provides more natural language and naturally occurring cohesion than simplified text. Simplified text is often criticized as creating unnatural discourse that reduces helpful redundancy and may, in effect, increase the reading difficulty of the text (Crandall, 1995). Supporters of simplified text, however, argue that beginning L2 learners benefit from text that is lexically, syntactically, and rhetorically less dense than authentic text. Despite the extensive dispute between the two camps, empirical evidence demonstrating the extent to which the two text-types differ has been rare (cf. Crossley et al., 2007). One goal of this paper is to empirically examine the validity of these two claims. To accomplish this, we analyze a corpus of both text-types in order to computationally compare their cohesive features. Additionally, we demonstrate an approach to successfully distinguishing the two text-types based on these same cohesive features.

Coh-Metrix

Recent advances in various disciplines have made it possible to computationally investigate measures of text and language comprehension that supercede surface components of language and instead explore deeper, more global discourse attributes (Graesser et al., 2004). *Coh-Metrix* is a computational tool that measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis. *Coh-Metrix* enhances conventional readability measures and provides detailed language and cohesion features (Graesser et al., 2004). The system integrates lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics (Jurafsky & Martin, 2000). *Coh-Metrix* analyzes texts on several dimensions of cohesion, including: *coreferential cohesion*, *causal cohesion*, *density of connectives*, *latent semantic analysis*, and *syntactic complexity*. For reasons of comparison, *Coh-Metrix* also includes standard readability measures such as Flesch-Kincaid Grade Level (Klare, 1974-1975) and several metrics of word and language characteristics such as word frequency, parts of speech, concreteness, polysemy, density of noun-phrases, and familiarity measures (Graesser et al., 2004).

Coh-Metrix has been used in a wide variety of studies including reading comprehension studies (Best, Ozuro, & McNamara, 2004) and text identification (Louwerse et al., 2004). *Coh-Metrix has* also proven to be effective at fleshing out differences between similar text-types. This includes internal structure analysis (McCarthy, Briner et al., 2006) and authorship attribution studies (McCarthy, Lewis, et al., 2006). In light of the past successes of *Coh-*

Metrix at distinguishing between texts and genres, the *Coh-Metrix* tool seems particularly well-suited for the purposes of this study.

Corpus Design

We constructed a corpus of second language texts, including both authentic and simplified text examples. In total, 224 texts used for second language instruction were excerpted from 11 intermediate L2 reading textbooks marketed for adult learners in second or foreign language learning environments (see Table 1). All texts in the selected readers of 100 words or more were included in the analysis. Text size was not considered a factor, particularly since *Coh-Metrix* normalizes its findings based on text length or provides a normalized ratio score.

Table 1: Corpus Information.

	Authentic	Simplified
Number of text books	4	7
Number of texts	101	123
Mean Number of words	696.37	471.23
Standard Deviation	555.71	222.39
Total Words in Corpus	70,333	57,961

Results

To examine the hypothesis that there are linguistic differences that differentiate simplified and authentic texts, we conducted a discriminant function analysis. A discriminant function analysis is a common approach used in many previous studies that attempt to distinguish between text-types (e.g., Biber 1993; McCarthy, Lewis, et al., 2006). We determined that five indices would be an appropriate number of predictors (one predictor for every twenty variables) that would not create problems of overfitting. As there are more than five indices made available through *Coh-Metrix*, we included only theoretically relevant variables. To this end, we divided our variables into three categories: *discourse* level, *sentence* level, and *word* level, following Graesser and Haberlandt (1985) and Vellutino (2003).

Based on the work of Crossley et al. (2007), these three categories were further organized into five distinct banks of indices: *lexical co-referentiality, logical connectors, syntactic complexity*, *textual abstractness/ambiguity*, and *word information*. We selected one index from each of the five identified banks based on the effect size of the difference between the two sets. To select the variable from the banks, we divided the dataset into a *training set* (*n*=113 texts) and a *test set* (*n*=111 texts). An ANOVA was conducted on each of the banks of variables in the training set, which selected the variable with the largest effect size as the representative variable of that bank (see Table 2). A discriminant function analysis, including the five variables with the largest effect size from each bank,

Table 2: Means, F-values, and Effect Sizes for each of the Five Categories Comparing Authentic and Simplified Texts.

Variables	Authentic	Simplified	$F(1,112)$	η^2
Noun Overlap	0.191 (0.130)	0.277 (0.189)	7.536*	0.064
Logical Connectors	47.963 (15.472)	35.242 (10.132)	27.569**	0.199
Syntactic Complexity	3.193 (0.650)	2.750 (0.764)	10.691**	0.088
Verb Hypernymy	1.917 (0.137)	1.859 (0.604)	4.212*	0.037
Age of Acquisition	328.039 (46.595)	312.565 (30.771)	4.472*	0.039

Note: standard deviations are in parentheses, *p <.05 ** p <.001

was conducted with text-type (authentic or simplified) as the dependent variable. The structure matrix with the coefficients for each function for each variable is shown in Table 3.

Table 3: Structure Matrix from Discriminant Analysis

	Authentic	Simplified
Noun Overlap	-33.632	-25.461
Logical Connectors	0.112	0.053
Syntactic Complexity	10.283	8.352
Verb Hypernymy	88.036	85.359
Age of Acquisition	0.236	0.221
Constant	-139.745	-123.448

Lexical Co-Referentiality. The variable selected to represent the lexical co-referentiality bank was *Noun Overlap, all distances, unweighted* (the term *unweighted* refers to the simple average overlap across each of the sentence pairs, without considering distance from the target sentence). Noun overlap measures how often a common noun exists between two sentences. These types of cohesive links have been shown to aid in text comprehension and reading speed (Kintsch & van Dijk, 1978). In simplified texts, co-referentiality is important because simplified texts are often created with considerations for increased clarification and elaboration (Young, 1999) and publishers' guidelines urge writers of simplified texts to take great care with pronominal reference (Simensen, 1987).

The results of the study suggest that simplified texts have greater levels of coreference cohesion. This result is similar to the finding of Crossley et al. (2007) and provides further evidence that simplified text, with its dependency on noun phrases, avoidance of pronominal reference, and simple syntactic structure, provides a greater amount of co-referentiality. As a result, in terms of co-reference, simplified text appears to show greater cohesion than authentic text. This is not surprising considering that simplified text provides smaller lexical domains than authentic text, which, in turn, allows for less lexical variance and more co-referential overlap. Thus, simplified text provides more redundancy, which, when available from more than one source, assists readers in understanding the message and intention of a text.

Additionally, redundancy allows readers to build connections between sections of text (Haber & Haber, 1981; Smith, 1988).

Logical Operator Incidence. The variable selected to represent the logical connector bank was *Logical Operator Incidence Score*, which tracks the incidences of *and's*, *if's*, *or's*, *conditional constructions*, and *negations*. The logical operators measured in *Coh-Metrix* include variants of *or*, *and*, *not*, and *if-then* combinations, all of which have been shown to relate directly to the density and abstractness of a text and correlate to higher demands on working memory (Costerman & Fayol, 1997). Logical operators are commonly simplified in second language reading texts as a result of publishers' guidelines that call for lexical control and the careful use of connectives in second language reading texts (Simensen, 1987).

The results of the study suggest that the authentic texts contain a greater distribution of logical operators. This result supports the earlier findings of Crossley et al. (2007). This result is not particularly surprising as logical connectors are considered a more complex feature of English syntax and should likely be more common in authentic text than in simplified text. Logical operators are important, of course, for discussing hypothetical situations and logically linking phrases. Their comparative absence from simplified text, which could be seen as providing a more concrete, less abstract text, also limits the discourse structure of the text. The lack of logical operators may lead to a less abstract text, but the use of logical operators has been identified as one of the major lexical features that constitute cohesive bonds between sections of text (Halliday, 1985) and the minimalization of such bonds in simplified text could lead to readings that do not elaborate, extend, and enhance the ideas of the text to the extent that an authentic text may.

Syntactic Complexity. The variable selected to represent the syntactic complexity bank was *Mean Number of Higher Level Constituents per Sentence*. *Coh-Metrix* measures syntactic complexity by measuring the mean number of higher level constituents, defined in *Coh-Metrix* as sentences and embedded sentence constituents, per noun phrase. According to *Coh-Metrix*, sentences with difficult syntactic composition have a higher ratio of constituents per word and noun phrase (Graesser et al., 2004). Variables such as this are important in the simplification of second language reading texts because simplified texts are

often simplified through the use of shorter sentences, reduced language features, and specified grammatical constructions (Long & Ross, 1993).

The results of the study offer evidence that authentic texts are more syntactically complex than simplified texts. Greater syntactic complexity in authentic texts supports the claim that simplified texts are more syntactically accessible to second language readers than authentic texts (Bamford, 1984). This result differs from that reported in Crossley et al. (2007) who found that simplified texts at the beginning level were more syntactically complex than authentic texts. However, the difference between the two studies is likely caused by the difference in corpora levels (beginning versus intermediate) as intermediate texts move toward more natural syntactic constructions and away from short sentences that rely on simple sentence constructions and elaboration. This intermediate text analysis contradicts many of the criticisms directed towards simplified text. While it was true that beginning simplified text likely contains atypical language structures and short, choppy syntax, it is unlikely that simplified text at the intermediate level is nearly as problematic or leads to text that is unnaturally plain.

Verb Hypernymy. The variable selected to represent the textual abstractness/ambiguity bank was *Mean Hypernym Values of Verbs. Coh-Metrix* measures hypernymy values, which refer to the number of levels a word has in a conceptual, taxonomic hierarchy, using WordNet (Fellbaum, 1998; Miller et al., 1990). The hypernymy value of verbs are often simplified in second language reading texts as a result of publishers' guidelines that call for the control of abstractness (Simensen, 1987).

The results of the study suggest that authentic texts are more abstract than simplified texts when other variables are taken into account. Such a result supports the findings of Crossley et al. (2007) and corroborates the claim that authentic texts are often more abstract than simplified texts and might therefore create a heavier comprehension burden on the second language reader (Bamford, 1984).

Age of Acquisition. The variable selected to represent the word information bank was *Age of Acquisition, Mean for Every Word. Coh-Metrix* measures lexical age of acquisition through the MRC Psycholinguistic Database (Coltheart, 1981). Lexical age of acquisition scores are important in the simplification of second language texts because much of the simplification process consists of trimming words and phrases and deleting low frequency vocabulary words (Young, 1999).

The results of the study offer evidence that simplified texts contain words that have a lower age of acquisition than authentic texts. This supports the findings of Crossley et al (2007) and strengthens the idea that the lexicon used in simplified text is more frequent and familiar and therefore more accessible to second language learners

(Bamford, 1984) as the use of more frequent words in simplified text should allow it to be more quickly processed by beginning readers and is thus an advantage for L2 readers with regard to automaticity and word recognition (Carrell & Grabe, 2002; Woodinsky & Nation, 1988).

Accuracy. An estimation of the accuracy of analysis can be made by plotting the correspondence between the actual text-type in the test and training sets and the predictions made by the discriminant analysis (see Table 4). The results show that the discriminant analysis correctly allocated 156 of the 224 texts, an average accuracy rate of 70% (df=1, N=224) x^2= 33.55, $p < .001$). However, this figure might be slightly inflated by the inclusion of the data from the training set. Using the test data set only, the accuracy of the discriminant analysis is somewhat lower, with 67 of the 110 texts in the testing set being analyzed correctly for an average accuracy rate of 60% (df=1, N=111) x^2= 4.55, p < .05). The precision, recall, and F1 measures for each text-type further demonstrate the accuracy of the model (see Table 5).

Table 4: Predicted text-type versus actual text-type featuring results from both training set and test set.

Actual Text-type	Predicted Text-type	
Training set	Authentic	Simplified
Authentic	**36**	15
Simplified	9	**53**
Test Set	Authentic	Simplified
Authentic	**29**	21
Simplified	23	**38**

Table 5: Precision and Recall Findings (Testing and Training Sets).

Training Set

Text Set	Precision	Recall	F1
Authentic	0.800	0.706	0.753
Simplified	0.779	0.855	0.817

Test Set

Text Set	Precision	Recall	F1
Authentic	0.558	0.580	0.569
Simplified	0.644	0.623	0.633

Discussion

In this study we analyzed a corpus of simplified and authentic second language reading texts using the computational tool *Coh-Metrix*. The purpose of the study was to examine whether a tool such as *Coh-Metrix* could

discriminate between comparable text-types and provide useful information about the subtle differences between texts. The results of this study suggest that computational tools such as *Coh-Metrix* can be used as a means of distinguishing groups of similar text-types. From a practical standpoint, the findings provide researchers interested in the field of second language material development with fundamental information about how simplified and authentic texts differ and to what degree. In general these results support the findings of Crossley et al. (2007) in demonstrating that authentic texts contain more logical connectors and have higher degrees of syntactic complexity, verb abstractness, and age of acquisition scores. In contrast, authentic texts tend to display lower levels of co-referentiality. These findings help to substantiate the idea that authentic text and simplified text differ in their use of language and that these differences have important consequences for the second language learner, teacher, and material developer. Most importantly, these findings further clarify past claims about the value of both simplified and authentic text in terms of how they can lead to the acquisition of a second language. These findings may also come to better inform classroom pedagogy by providing both classroom teachers and material developers with crucial information about the linguistic features found in simplified and authentic text. This information will allow materials developers to better understand how the process of simplification modifies the linguistic structure of text and how these modifications might interact with second language reading processes and comprehension. These findings will allow for the development of simplified text and the selection of authentic text that are more comprehensible for second language readers. Classroom teachers, in a like manner, will have additional information about the linguistic differences between authentic and simplified text and will be able to self-select text that will better match the needs of their students.

Language production is a complex process that in many ways, especially at the discourse level, is still poorly understood. Artificially modifying texts according to a few simple pedagogical principles can cause unintended consequences for the overall structure of the discourse, potentially affecting how the text is processed, comprehended, and understood. In the case of second language learners, this may have important ramifications. The recognition of how these modifications interact with second language processing may allow for important change to be made in second language reading text such as allowing for the development of simplified texts that include more logical connectors to provide beneficial and natural links between ideas or the selection of authentic texts that provide more co-referentiality and are less syntactically complex. Progress such as this would allow for more readable texts and greater access to comprehensible input.

While providing important information about the linguistic differences between simplified and authentic text, this study also supplies evidence useful to researchers in the fields of data retrieval, text identification, natural language processing, and forensic linguistics. The better able we are to distinguish between seemingly similar texts the better able we will be to retrieve documents accurately, to assign more likely syntax structures to parsers, and to identify texts that may be fraudulent.

Our future research will concentrate on creating larger and more diverse corpora in order to confirm and extend the results of this study. We are particularly interested in whether general differences between authentic and simplified texts extend to differences between beginner and intermediate learning texts. While much work remains to be done, the information presented in this study contributes to a variety of fields by providing evidence for approaches to text-type identification, evidence of textual differences for material developers in second language learning, and evidence that even shallow, moderate textual changes can significantly affect discourse structures.

Acknowledgments

References

Allen, J. & Widdowson, H.G. 1979. Teaching the communicative use of English. In Brumfit, C. & Johnson, K. Eds. *The Communicative Approach to Language Teaching*. Oxford: Oxford University Press.

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

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. 2005. Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, 25, 65-83.

Biber, D. 1993. Representativeness in corpus design. *Journal of Literary and Linguistic Computing,* 84, 243-257.

Carrell, P., & Grabe, W. 2002. Reading. In N. Schmitt Ed., *An introduction to applied linguistics*. New York: Oxford University Press.

Coltheart, M. 1981. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Colwell, K., Hiscock, C.K., & Memon, A. 2002. *Interviewing techniques and the assessment of statement credibility. Applied Cognitive Psychology, 16*, 287-300.

Costerman, J. & Fayol, M. 1997. Processing interclausal relationships: Studies in production and comprehension of text. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cowan, J.R. 1976. Reading. Perceptual strategies, and contractive analysis. *Language Learning*, *26*, 95-109.

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*, New York: Heinle & Heinle Publishers.

Crossley, S.C., Louwerse, M.L, McCarthy, P.M., McNamara, D.S. 2007 What is an authentic text? A computational analysis of second language reading texts. *Modern Language Journal, 91 2.*

Davies, A., & Widdowson, H. 1974. Reading and writing. In J. P. Allen & S. P. Corder Eds. *Techniques in applied linguistics*, Oxford: Oxford University Press.

Goodman, K. 1976. Psycholinguistic universals in the reading process. In P. Pimsleur and T. Quinn Eds. *The psychology of second language learning*, Cambridge, UK: Cambridge University Press.

Graesser, A.C., McNamara, D.S., & Louwerse, M.M. 2003. What do readers need to learn in order to process coherence relations in narrative and expository text. In A. P. Sweet and C. E. Snow Eds., *Rethinking reading comprehension*. New York: Guilford Publications.

Graesser, A., McNamara, D.S., Louwerse, M.M., & Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 193-202.

Haber, R. N., & Haber, L. R. 1981. Visual components of the reading process. *Visible Language, 15*, 147-182.

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

Halliday, M.A.K. 1985. *An introduction to functional grammar*. Baltimore: Edward Arnold.

Honeyfield, J. 1977. Simplification. *TESOL Quarterly*, *11*, 431-440.

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

Johnson, K. 2001. The communicative approach to language Teaching. Oxford: Oxford University Press.

Jurafsky, D., & Martin, J.H. 2000. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, Upper Saddle River, NJ: Prentice-Hall.

Kao, A. & Poteet, S. 2006. *Natural Language Processing and Text Mining*. London: Springer-Verlag UK.

Kessler, A., Nunberg, G., & Schutze, H. 1997. Automatic detection of text genre. *Proceedings of the 35th Annual Meeting of Association for Computational Linguistics*, and *The 8th Conference of European Chapter of Association for Computational Linguistics,* Madrid, Spain.

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

Klare, G.R. 1974-1975. Assessing readability. *Reading Research Quarterly*, 10, 62-102.

Kuo, C. 1993. Problematic issues in EST materials development. *English for Specific Purposes*, *12*, 171-181.

Little, D., Devitt, S. & Singleton, D. 1989. *Learning foreign languages from authentic texts: Theory and practice*. Dublin, Ireland: Authentik.

Long, M. & Ross, S. 1993 Modifications that Preserve Language and Content. in M. L. Tickoo Ed. *Simplification: Theory and application*, Singapore: SEAMEO Regional Language Center.

Louwerse, M.M., McCarthy, P.M., McNamara, D.S., & Graesser, A.C. 2004. Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, T. Regier Eds. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, Mahwah, NJ: Erlbaum.

Mackay, R. 1979. Teaching the information gathering skills. In R. B. Barkman & R. R. Jordan Eds. *Reading in a Second Language*, Mackay, Rowley, Mass: Newbury House.

McCarthy, P.M, Briner, S.W, Rus, V., & McNamara, D.S. 2006. Textual Signatures: Identifying text-types using Latent Semantic Analysis to measure the cohesion of text structures. In: A. Kao, & S. Poteet Eds.. *Natural Language Processing and Text Mining*. London: Springer-Verlag UK.

McCarthy, P.M, Lewis, G.A., Dufty, D.F., & McNamara, D.S. 2006. Analyzing writing styles with Coh-Metrix. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference FLAIRS,* Melbourne, Florida.

Miller, G.A., 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–312.

Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. 2003. Lying words: Predicting deception from linguistic style. *Personality and Social Psychology Bulletin*, *29*, 665–675.

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

Smith, F. 1988. *Understanding reading*, 4th ed. Hillsdale, New Jersey: Erlbaum.

Vellutino, F.R. 2003. Individual differences as sources of variability in reading comprehension in elementary school children. In A. P. Sweet & C. E. Snow Eds., *Rethinking reading comprehension,* New York: Guilford Press.

Wodinsky, M., & Nation, P. 1988. Learning from graded readers. *Reading in Foreign Languages*, *5*, 155-161.

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