# Using *Coh-Metrix* to Assess Differences between English Language Varieties.

Charles Hall*, Philip M. McCarthy, Gwyneth A. Lewis, Debra S. Lee**, and
Danielle S. McNamara
Department of Psychology, University of Memphis
* Department of English, University of Memphis
** CEELI Institute, the Czech Republic

This study examined differences between the written, national language varieties of the United States and Great Britain, specifically in texts regarding the topic of Law. The few previous studies that have dealt with differences between the dialects of the United States and Great Britain have focused on shallow-level features, such as lexis, subject-verb agreement, and even orthography. In contrast, this study uses the computational tool, *Coh-Metrix*, to distinguish British from American discourse features within one highly similar genre, Anglo-American legal cases. We conducted a discriminant function analysis along five indices of cohesion on a specially constructed corpus to show those differences in over 400 American and English/Welsh legal cases. Our results suggest substantial differences between the language varieties, casting doubt on previous generalizations about British and American writing that predict that the national varieties would vary more by genre than by language variety. Our results also offer guidance to materials developers of legal English for international purposes (such as in the E.U.) and drafters of international legal documents for producing effective and appropriate materials.

## 1. Introduction

George Bernard Shaw once said: "England and America are two countries divided by a common language." A century later, there has been relatively little work done to assess what those divisions may be (cf., Biber, 1987; Helt, 2001). In this paper, we add to previous research by demonstrating a computational model capable of distinguishing significant differences between a text type that has traditionally been thought to be highly similar in the American and British national language varieties: *Legal English*. We argue that the success of our study provides the necessary foundation for future research assessing the linguistic distinctions between these English language varieties. We also argue that our results offer evidence that other specific English language registers may also differ significantly depending on language variety. Finally, we argue that our results offer substantial guidance to materials developers in the domain of *legal English for international purposes* (such as in the European Union) and for

drafters of international legal documents toward producing effective and appropriate materials.

## 2.      Purpose of the Research

The common assumption that greater differences exist along the lines of English *genres* (e.g., expository texts/narrative texts) than along the lines of *language varieties* (standard written British or American) within a specific written genre (Johansson, 1985) has helped to restrict the number of studies that compare American and British varieties of English. The few studies that have been done tend to only highlight well-established aspects at the phonological, lexical, or morphological level (Biber, 1987; Helt, 2001). This relative lack of language variety distinction studies highlights the two primary purposes of this paper.

The first purpose of this study is to provide researchers with a methodology for the computational distinction of highly similar text-types. Growing technology has provided an increased interest in such fields as *text mining* and *question answering software*. Such research, however, requires the identification of the subtlest of differences between relatively similar texts (e.g., McCarthy, Briner, Rus, & McNamara, 2007; McCarthy, Lewis, Dufty, & McNamara, 2006; Rus, McCarthy, & Graesser, 2006). The computational system and methodological approach outlined in this paper serve to facilitate such research by demonstrating that highly similar text types are computationally distinguishable by language variety alone.

The second purpose for this study reflects our choice to conduct a corpus analysis of register varieties within the specialized domain of *law*. Corpus analyses are recognized by the field of English for Specific Purposes (ESP) as being a primary source for helping materials developers produce more effective, appropriate texts (Biber, Conrad, & Reppen, 1998), which in turn serve to help students acquire the most important communicative skill in ESP, genre awareness (Bhatia, 1997). While genre data have grown steadily in traditional ESP areas (e.g., science, engineering, and business), no large-scale corpus study of law texts has previously been conducted.

Such a corpus is of particular importance to what Hall and Lee (2003) term "non-equivalent" domains of ESP. "Equivalent" fields, such as engineering or mathematics, work with basically the same content and approaches regardless of language or language variety. In contrast, domains such as *law* or *education* are "non-equivalent" since their field-specific content and presuppositions vary according to the local system, not language.

In this study, therefore, we examine a specially created corpus of British and American legal texts in order to provide empirical evidence to establish differences

between the language varieties that will tend to also support the assertion that the construct often named Anglo-American law is actually composed of two mildly non-equivalent domains. As such, this study establishes a technique through which material writers in ESP may analyze and assess specialized corpora; it also serves to provide information of significant importance to researchers in computational fields such as text mining and question answering systems where distinguishing highly similar text types is paramount.

## 3.    Coh-Metrix

The first multi-dimensional study across English language varieties was conducted by Biber (1987). Biber contrasted nine written genres, finding evidence that British texts are more formal but less interactive and abstract than their American counterparts. While Biber noted the differences were not large, he maintained that they nonetheless remained consistent, forming an underlying linguistic difference between the varieties with greater grammatical and stylistic prescription being present in the British variety. These findings were supported by Helt (2001) who updated Biber's approach and found similar differences within the *spoken* registers.

The shallow metric approach to distinguishing varieties of English used by both Biber and Helt has certainly been profitable. However, recent developments in computational linguistics and discourse processing have made it possible for researchers to develop indices that are far more sophisticated. These indices have been gathered together in a tool called *Coh-Metrix* (Graesser, McNamara, Louwerse, & Cai, 2004) developed at the University of Memphis.

Coh-Metrix can process over 400 indices of cohesion, language, and readability, which together allow the tool to estimate a wide range of textual features reflecting cohesion relations, and world knowledge, together with language and discourse characteristics. Coh-Metrix functions through a variety of modules including syntactic parsers (Charniak 2000); latent semantic analysis (LSA, Landauer, Foltz, & Laham, 1998, Landauer, McNamara, Dennis, & Kintsch, 2006), and many other computational linguistics features (Jurafsky & Martin 2000). In addition to its sophisticated indices, Coh-Metrix also provides researchers with a range of traditional textual measures such as *average word length*, *average sentence length*, and the readability formulas of *Flesch Reading Ease* and *Flesch-Kincaid Grade Level* (Klare 1974-1975).

Since its inception in 2002, Coh-Metrix has been used to help establish a wealth of evidence on a variety of textual differences. For example, McCarthy, Lewis et al. (2006) demonstrated that Coh-Metrix was an effective tool in detecting authorship even when individual authors recorded significant shifts in their writing

style. McCarthy et al. (2007) used Coh-Metrix LSA indices to demonstrate structural cohesion across variously themed psychology articles. Duran, McCarthy, Graesser, and McNamara (2006) used Coh-Metrix to assess temporal coherence across the textual domains of *narratives*, *history*, and *science*. Other Coh-Metrix studies include distinguishing high and low cohesion texts (McNamara, Ozuru, Graesser, & Louwerse, 2007); estimating human assigned grade levels of published text books (Dufty, Graesser, Louwerse, & McNamara, 2006); calculating textual genre (Duran & McNamara, 2006; McCarthy, Graesser, & McNamara, 2006); assessments of the structural organization of published high school text books (Lightman, McCarthy, Dufty, & McNamara, 2006a; Lightman, McCarthy, Dufty, & McNamara, 2006b); assessments of formal/informal and spoken/written distinctions across genres (Louwerse, McCarthy, McNamara, & Graesser, 2004; Dempsy, McCarthy, & McNamara, 2006); studies of gender differences across texts (Bell, McCarthy, & McNamara, 2006); and assessments of authentic and modified texts published for students of English as a second language (Crossley, Louwerse, McCarthy, & McNamara, 2007; Crossley, McCarthy, & McNamara, 2006).

This wide variety and wealth of successful studies provides ample evidence that Coh-Metrix is an ideal tool for investigating differences across closely related registers. In the remainder of this section, we present a brief summary of the key Coh-Metrix banks of indices that make such analyses possible. For the reader interested in an extensive analysis of Coh-Metrix theory, modules and measures, Graesser et al. (2004) is recommended.

## 3.1 Causal cohesion

Causal cohesion compares the number of semantically identifiable causal verbs (e.g. *kill*, *drop*, *fill*) to the number of causal particles and semantically depleted verbs (e.g., *because*, *in order to*, *cause*, *make*). Causal cohesion has been found to have an important effect on improving text comprehension and recall (Keenan, Baillet, & Brown, 1987).

## 3.2 Coreferential cohesion

Referential links aid textual comprehension by facilitating inferencing and recall (Kintsch & van Dijk, 1978; McNamara, 2001). *Noun overlap*, *argument overlap*, *stem overlap*, and LSA-based semantic overlap are used by Coh-Metrix to identify lexical coreference. Lexically based pairs such as *chair*/*chairs* and *walk*/*walking* are compared in order to attain overlap scores. LSA-based measures analyze the relationship between textual elements on a semantic level. Thus, Coh-Metrix is

capable of identifying and assessing semantically similar pairs such as (*finger/ hand* and *window/ glass)*.

## 3.3    Connectives and logical operators

Coh-Metrix measures the density of connectives (the cohesive links between separated conceptual units) using categories such as positive-additive connectives (e.g., *additionally*), negative-additive connectives (e.g. *conversely*, *instead*), positive-temporal connectives (*after this*, *earlier*, *finally*), and negative temporal connectives (e.g. *until*). These various connectives are considered vital indicators of cohesion (Halliday & Hasan, 1976; Louwerse, 2002; Graesser et al., 2004). The density of logical operators (e.g. *or*, *and)* is also measured as these items contribute to the level of working memory readers require for comprehension (Graesser et al., 2004).

## 3.4    Density of major parts of speech

Coh-Metrix offers density scores for various parts of speech (POS), including pronouns, nouns, verbs, adjectives, adverbs, cardinal numbers, determiners, and possessives. POS are given by the Brill (1995) POS tagger and are reported by their incidence per 1000 words in the text. The density of such POS is an indicator of textual difficulty since, for example, more pronouns may result in a greater cognitive strain on the reader (Graesser et al., 2004).

## 3.5    Polysemy and hypernymy

WordNet (Fellbaum, 1998) is used to calculate values for lexical polysemy (the number of senses or meanings a word has), and hypernymy (the number of levels in a conceptual, taxonomic hierarchy). Such indices are used by Coh-Metrix to indicate the potential ambiguity and/or abstractness of a text.

## 3.6    Syntactic complexity

Syntactic complexity reflects the degree to which sentential clauses and phrases are embedded in the text. Greater complexity reflects greater textual ambiguity, structural density, and ungrammaticality (Graesser et al., 2004).

### 3.7 Word information and frequency

Coh-Metrix incorporates four banks of indices (*familiarity*, *imageability*, *concreteness*, and *meaningfulness*) to provide values for word information. These scores are derived from the MRC Psycholinguistic database (Coltheart, 1981). Such values help in the assessment of texts as, for example, high frequency words are more easily understood because they are more common to the reader, resulting in faster reading times (Haberlandt & Graesser, 1985; Just & Carpenter, 1980).

### 3.8 Other indices

Shallow indices (e.g. *syllable count*, *word length*, *sentence length*, *number of words per sentence/paragraph/text)* are the traditional indices of text assessment. Such indices in various combinations provide measures such as Flesch Reading Ease and Flesch-Kincaid Grade Level (Klare, 1974-1975). Coh-Metrix provides these measures and a host of other similar shallow indices. In the current study, only the newer more interesting of these measures were included as many traditional indices have previously been used in stylistic analysis (e.g. Brinegar, 1963; Fucks, 1952), Such indices are important for detecting mode, genre, and style in text; however, our focus in this study is in determining the role of cohesion in language varieties.

### 4. Corpus and Method

To represent the register of law, we selected texts from *commercial competition* cases. This choice of comparison text type was made for a variety of reasons. First, competition cases are plentiful in both American (US) and English/Welsh[1] (EW) law. Second, access to e-texts of these cases is provided through large online databases[2]. Third, competition cases typically provide long, detailed texts that are suitable for an initial Coh-Metrix analysis.

To locate US cases within the databases, we used the key words *antitrust* and *competition*. Since the term *antitrust* is not used frequently in EW law, we used the key words *competition* and *commercial* to locate a similar EW range of cases. All cases were selected for recency, dating no earlier than 1991. Each case was also manually checked to ensure that it was a genuine competition case. With false cases removed, the corpus contained 408 cases (US = 200, EW = 208).

---

[1] *Scottish* cases were not used because they are based on a different legal system.
[2] http://www.bailii.org/form/search1.html and lawschool.westlaw.com

Finally, to normalize for text length (e.g., Biber, 1988; Louwerse et al., 2004), continuous sections with minimum lengths of 1000 words were randomly selected. All files were then processed through Coh-Metrix.

## 5.    Discriminant Analysis

To determine differences between the language varieties, we conducted a discriminant function analysis using the values produced by Coh-Metrix indices. Discriminant function analysis has been successfully used to distinguish text types in a variety of previous studies (Biber 1993; Crossley, McCarthy et al., 2006; Karlgren & Cutting 1994; McCarthy, Lewis et al., 2006). A discriminant function analysis produces a weighted linear equation for each category in an analysis. In the current study, the categories are *English/Welsh texts* and *American texts*. To establish the degree of success of such an analysis, the corpus in question is typically divided into two approximately equal halves: a *training set*, and a *test set*. The training set is used to generate coefficients via the discriminant analysis, and the test set is used to establish the degree of accuracy of those coefficients in predicting known cases.

To avoid problems of overfitting respective to the size of the dataset, we limited our analysis of the training set to five predictor indices. Overfitting is attempting to use too many variables for the number of observations in the data set. When overfitting occurs, unwanted *noise* (normal variation in different samples) is returned in addition to the *signal* of the predicators that the researcher wants. Since overfitting measures the noise of the training set, the overfitted training model appears to fit the data well but would lack accuracy when applied to new data (such as the testing set) since by definition noise differs from set to set.

Since there are many more than five indices made available through Coh-Metrix, we elected to focus this initial study on the more theoretically relevant and interesting indices. To this end, we following such studies as Haberlandt and Graesser (1985) and Vellutino (2003) and divided Coh-Metrix indices into three categories: *word*, *sentence*, and *discourse*. Word-level variables assess values of individual words in the text independent of other features of the text. For example, many word frequency indices simply report the frequency of a word's use independent of its use in the text under analysis. Sentence-level variables differ from word-level variables inasmuch as the context of the sentence is assessed in the metric. Thus, an index such as *number of words before main verb*, considers each sentence in the text and derives a mean score. At the discourse level, all sentences in the text become inter-dependent. For example, some cohesion indices compare each sentence in the text to every other sentence in the text. In such a way, the entire discourse is evaluated. In this study, we excluded the first two

*shallow* categories from the study as their previous use in similar analyses had yielded only limited success (e.g., Biber, 1987; Karlgren & Cutting, 1994). This left us with approximately 50 cohesion indices at the more sophisticated *discourse level*. These discourse indices bring with them a strong history of successful analysis, having already contributed to significant discriminant findings (e.g., Crossley, McCarthy, et al., 2006; Louwerse et al., 2004; McCarthy, Graesser et al., 2006).

These discourse indices comprise five distinct categories each containing many indices: *coreferential cohesion*, *casual cohesion*, *local-grammatical cohesion*, *latent semantic analysis* (LSA), and *lexical diversity* (discussed in greater detail below). We elected to include one predictor index from each of the five categories identified above. To do so, having randomly divided the dataset into a *training set* (n=200 texts), and a *test set* (n=208 texts), we conducted an analysis of variance (ANOVA) on the training set to select the variable with the largest effect size as the representative variable from each of the five categories. Table 1 presents the means, F values, and effect sizes for each of the five indices.

Table 1
Means, F values, and Effect Sizes for Co-referential Cohesion (co-ref cohesion, Causal Cohesion, Local-grammatical Cohesion (LG cohesion), Latent Semantic Analysis (LSA), and Lexical Diversity for Comparing English/Welsh (EW) and American (US) Legal Cases

|  | EW | US | $F(1,199)*$ | $\eta^2$ |
|---|---|---|---|---|
| Co-ref cohesion | 0.478 (0.133) | 0.267 (0.094) | 168.527 | 0.460 |
| Causal cohesion | 31.526 (7.640) | 26.383 (7.073) | 24.400 | 0.110 |
| LG cohesion | 41.243 (9.179) | 34.821 (8.730) | 25.699 | 0.115 |
| LSA | 0.194 (0.064) | 0.120 (0.034) | 125.253 | 0.392 |
| Lexical diversity | 0.562 (0.065) | 0.613 (0.058) | 33.934 | 0.146 |

Note: standard deviations are in parentheses, * *all values significant at p < .001.*

A discriminant function analysis was then conducted on the training set with language variety (English/Welsh or American) as the dependent variable. The structure matrix with the coefficients for each function for each variable is shown in Table 2.

Table 2
Structure of the Discriminant Functions for Predictors and Constant

| Variable | Domain | |
|---|---|---|
| | EW | US |
| Co-referential Cohesion | 22.441 | 11.115 |
| Causal Cohesion | 0.499 | 0.460 |
| Local-grammatical Cohesion | 0.597 | 0.533 |
| Latent Semantic Analysis | 40.948 | 25.374 |
| Lexical Diversity | 162.080 | 173.049 |
| (Constant) | -76.186 | -72.133 |

## 5.1 Co-referential cohesion

The index resulting in the largest difference between language variety within the category of co-referential cohesion was *unweighted argument overlap including the prior three sentences* (co-reference). The term *unweighted* refers to the simple average overlap across each of the sentence pairs, without considering distance from the target sentence. Argument overlap refers to overlap between nouns and between pronouns (e.g., *dog/dog*; *pen/pens*; *he/he*).

The results suggest that EW texts have greater levels of co-referential cohesion when the other variables in the analysis are also taken into account. Texts with greater levels of co-reference contain greater repetition of lexical arguments. Co-reference facilitates readers by reducing the inferences necessary to make connections between sentences (e.g., McNamara, 2001).

## 5.2 Causal cohesion

Among *causal cohesion* variables, *incidence of positive causal connectives* (causal) showed the largest difference between language varieties and thus was included in the discriminant analysis. Causal connectives (e.g., *since*, *so that*, *because*) explicitly connect ideas for the reader and facilitate comprehension and are more frequent in texts where causal events are prevalent such as scientific texts or narratives with action plots (Graesser, Singer, & Trabasso, 1994; Trabasso & van den Broek, 1985). The results indicate that the EW texts contain a greater distribution of causal connectives. Since legal cases often contain narrative-like restatements of the history of the case, this is an important distinction.

## 5.3    Grammatical-local cohesion

The variable chosen for the grammatical-local category was *incidence of logical operators* (local-grammatical). These elements (e.g., *or*, *and*, *not*, *if/then*) are more frequent in the EW than the US texts, and are indicative of an analytically dense text, which can place a high burden on working memory (Graesser et al., 2004).

## 5.4    Latent Semantic Analysis (LSA)

The greatest difference between language varieties among LSA indices was *sentence-to-sentence mean* (LSA). In this index, the texts are the adjacent sentences within paragraphs of the corpora. LSA differs from coreference indices in that it goes beyond lexical similarities such as *chair/chairs* and is able to rate the relative semantic similarity between terms such as *chair/table*, *table/wood*, and *wood/forest*. As such, LSA is able to evaluate not only whether two items are the same, but also the degree to which they are similar (McCarthy et al., 2007). The results offer evidence that EW texts are not only more explicitly cohesive, but are also more cohesive at an implicit, semantic level.

## 5.5    Lexical diversity

The variable yielding the largest difference between language varieties in the lexical diversity category was *type-token ratio for content words other than nouns* (TTR). Coh-Metrix incorporates TTR as its primary index of lexical diversity. When texts are of similar length (as in this study) TTR is a reliable index of the repetition of lexical items within a text (McCarthy & Jarvis, in press). Low TTR scores indicate greater repetition of lexical items, which should ease textual processing (Graesser et al., 2004). The results indicate that EW texts are comprised of a lower degree of lexical diversity. Biber (1987) used a very similar index of lexical diversity but found no evidence of a significant effect in his more general analysis of genres.

## 6.    Accuracy

The accuracy of the analysis can be best estimated by plotting the correspondence between the language varieties in the test set and the predictions made by the discriminant analysis. Since we know the actual composition of the 208 texts in test set, we can determine the accuracy of the predictions made by the discriminant analysis (see Table 3). In this study, as is typical of discriminant analyses' studies

(e.g. McCarthy, Lewis et al. 2006), the accuracy of the results are reported in terms of *recall, precision, and F1*. Recall shows the number of correct predictions divided by the number of true items in the group. In other words, recall is the number of *hits* over the number of *hits + misses*. Precision, on the other hand, is the number of correct predictions divided by the number of incorrect predictions. In other words, precision is the number of *hits* divided by the number of *hits + false alarms*. The distinction is important because an algorithm that predicts everything to be a member of a single group will account for all members of that particular group (scoring 100% in terms of recall) but will also falsely claim many members of other group(s), thereby scoring poorly in terms of precision. The F1 index provides an estimate of the combination of both recall and precision. Reporting each of these values allows for a better understanding of the accuracy of the model. The results show that the discriminant analysis correctly categorized 177 of the 208 texts, with an average accuracy rate of 85%. The precision, recall, and F1 measures for each language variety further demonstrate the accuracy of the model (see Table 4).

Table 3
Predicted Language Variety versus Actual Language Variety Applying the Model to the Test Set.

| language variety | Predicted language variety | |
|---|---|---|
| | EW | US |
| EW | **91** | 17 |
| US | 14 | **86** |

Table 4
Precision, Recall, and F1 Measures for each Language Variety.

| language variety | Precision | Recall | F1 |
|---|---|---|---|
| EW | 0.867 | 0.843 | 0.854 |
| US | 0.835 | 0.860 | 0.847 |

## 7.    Discussion

This study used a discriminant function analysis and five Coh-Metrix discourse-level predictor variables to investigate textual differences between English/Welsh

and US legal cases. Overall, the results indicate that the English/Welsh cases were more cohesive than were the US cases. The results also contradict the assumption made by Johansson (1985) that similar genres would vary little along language variety lines. The results provide evidence that Coh-Metrix is able to differentiate beyond diverse genres and modes and into closely related registers. The results are of special importance for materials developers in the rapidly growing field of ESP-law as the analyses help to determine which aspects of which language variety to teach in the growing international market. The information may also be of benefit in text mining as the analyses suggest that American and British English can be accurately distinguished. Specifically, the analysis indicates that given any one American or British law case on competition, the discriminant analysis function developed here would have an 85 percent likelihood of correctly identifying its country of origin.

Since this study used a relatively narrow field of register (antitrust/competition), our findings do not necessarily generalize to all legal areas. As such, future research will compare more registers (such as *legislation*, *court transcripts*, and "*boiler-plate*" documents) to determine if there are underlying, consistent differences in the British and American languages of law. Future analysis must also consider to what degree these differences in language variety extend to other genres such as narrative and expository.

The results of this study suggest that composition types may differ significantly depending on the community authorizing the language system rather than simply the language in which that text is composed. Consequently, we may have to consider the composition and comprehension of text beyond the genre level and into the language variety. To assess this possible development, future research may have to consider the degree to which expert human raters can distinguish differences between such language varieties. Consequently, we will need to explore further language varieties as a criterion in the composition and comprehension of text.

This initial study, offers compelling evidence that significant differences between language varieties do exist and can be computationally distinguished by a system such as Coh-Metrix. The algorithm generated in this analysis requires no human intervention at a judgment level and establishes that discourse level features are sufficiently diverse for sophisticated computational systems to distinguish texts with a very high degree of accuracy.

## Acknowledgements

expressed in this material are those of the authors and do not necessarily reflect the views of the IES. The authors would also like to thank Arthur Graesser, Max Louwerse, Zhiqiang Cai, and Yan Yan for their contributions to this project.

**References**

Bell, Courtney M., McCarthy, Philip M., and McNamara, Danielle S., 2006, July. *Variations in language use across gender: Biological versus sociological theories.* Poster presented at the 16th Annual Meeting of the Society for Text and Discourse, Minneapolis, MN.

Bhatia, V. 1997. Applied genre analysis and ESP. In *Functional Approaches to Written Text: Classroom Applications*, ed. Tom Miller. Washington, DC: USIA.

Biber, Douglas. 1987. A textual comparison of British and American writing. *American Speech*, *62*, 99-119.

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

Biber, Douglas, and Conrad, Susan. 2001. Corpus based research in TESOL. *TESOL Quarterly*, *35*, 331-335.

Biber, Douglas, Conrad, Susan, and Reppen, Randi. 1998. *Corpus linguistics*. Cambridge: Cambridge University Press.

Brinegar, Carl S. 1963. Mark Twain and the Quintius Curtius Snodgrass letters: A statistical test of authorship. *Journal of the American Statistical Association*, *58*, 85-96.

Charniak, Eugene. 2000. A maximum-entropy-inspired parser. *Proceedings of the First Conference on North American Chapter of the Association For Computational Linguistics* pp. 132-139. San Francisco, CA: Morgan Kaufmann Publishers.

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

Crossley, Scott A., Louwerse, Max M., McCarthy, Philip M., and McNamara, Danielle S. 2007. What is an authentic text: A computational analysis of second language reading texts. *Modern Language Journal*.

Crossley, Scott A., McCarthy, Philip M., and McNamara, Danielle S. 2006. Detecting manipulated texts. *Submitted to FLAIRS 2007*.

Dempsey, Kyle B., McCarthy, Philip M., and McNamara, Danielle S., 2006. Identifying text genres using phrasal verbs. *Submitted to FLAIRS 2007*.

Dufty, David F., Graesser, Arthur C., Louwerse, Max M., and McNamara, Danielle S., 2006. Is it just readability, or does cohesion play a role? *Proceedings of the 28th annual conference of the Cognitive Science Society.*Vancouver, Canada.

Duran, Nicholas D., McCarthy, Philip M., Graesser, Arthur C., and McNamara, Danielle S. 2006. An empirical study of temporal indices. *Proceedings of the 28th annual conference of the Cognitive Science Society.*Vancouver, Canada.

Duran, Nicholas D, and McNamara, Danielle S. 2006, July. *It's about time: Discriminating differences in temporality between genres.* Poster presented at the 16th Annual Meeting of the Society for Text and Discourse, Minneapolis, MN.

Fellbaum, Christiane. 1998. *WordNet: An electronic lexical database*. Cambridge: MIT Press.

Fucks, Wilhelm. 1952. On the mathematical analysis of style. *Biometrica*, *39*, 122-129.

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

Graesser, Arthur. C., Singer, Murray., and Trabasso, Tom. 1994. Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371-95.

Haberlandt. Karl, and Graesser, Arthur. 1985. Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, *114*, 357-374.

Haliday, Michael A., and Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman

Hall, Charles, and Lee, Debra S. 2003. What law to teach plenary address. *International Global English for Global Under-standing Conference*, Moscow, Russia.

Helt, Marie. 2001. A multi-dimensional comparison of British and American spoken English. In *Variation in English: Multi-dimensional studies*, eds. Susan Conrad and Douglas Biber. Essex: Pearson.

Johansson, Stig. 1985. Some observations on word frequencies in three corpora of present-day English texts. *I.T.L. Review of Applied Linguistics*, (*67-68*), 117-26.

Jurafsky, Daniel, and Martin, James. 2000. *Speech and language processing*. Englewood, NJ: Prentice Hall.

Just, Marcel A., and Carpenter, Patricia A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-354.

Karlgren, Jussi, and Cutting, Douglas. 1994. Recognizing text genres with simple metrics using discriminant analysis. *Proceedings of the 15th conference on Computational linguistics*. Morristown, NJ: Association for Computational Linguistics.

Kintsch, Walter, and van Dijk, Teun A. 1978. Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Keenan, Janice M., Baillet, Susan D., and Brown, Polly. 1987. The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*, 115-126.

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

Landauer, Thomas K., Foltz, Peter W., and Laham, Darrel. 1998. An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Landauer, Thomas., McNamara, Danielle., Dennis, Simon, and Kintsch, Walter. eds. 2006. *LSA: A road to meaning*. Mahwah, NJ: Erlbaum.

Lightman, Erin J., McCarthy, Philip M., Dufty, David F., and McNamara, Danielle S. 2006a. *Using Coh-Metrix to assess the structural organization of narratives*. Paper to be presented at the Society for Text and Discourse conference, Minneapolis, MN.

Lightman, Erin J., McCarthy, Philip M., Dufty, David F., and McNamara, Danielle S. 2006b. Using Coh-Metrix to assess distributions of cohesion and difficulty in high-school textbooks. *Submitted to FLAIRS 2007*.

Louwerse, Max M. 2002. An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics*, *12*, 291-315.

Louwerse, Max M., McCarthy, Philip M., McNamara, Danielle S., and Graesser, Arthur. 2004. Variation in language and cohesion across written and spoken registers. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

McCarthy, Philip M., Briner, Stephen W., Rus, Vasile., and McNamara, Danielle S. 2007. Textual Signatures: Identifying text-types using Latent Semantic Analysis to measure the cohesion of text structures. In *Natural Language Processing and Text Mining*, eds. Ann Kao and Steve Poteet. London: Springer-Verlag UK.

McCarthy, Philip M., Graesser, Arthur C., and McNamara, Danielle S. 2006, July. *Distinguishing genre using Coh-Metrix indices of cohesion.* Paper presented at the Society for Text and Discourse conference, Minneapolis, MN.

McCarthy, Philip. M., and Jarvis, Scott (in press). A theoretical and empirical evaluation of *vocd*. *Language Testing*.

McCarthy, Philip M, Lewis, Gwyneth A, Dufty, David F., and McNamara, Danielle S. 2006. Analyzing writing styles with Coh-Metrix. *Proceedings of the Florida Artificial Intelligence Research Society International Conference FLAIRS*, Melbourne, Florida.

McNamara, Danielle S. 2001. Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51-62.

McNamara, Danielle S., Ozuru, Yasuhiro., Graesser, Arthur C., and Louwerse, Max. 2006. Validating Coh-Metrix. *Proceedings of the 28th annual conference of the Cognitive Science Society*. Vancouver, Canada.

Rus, Vasile, McCarthy, Philip M., Graesser, Arthur C. 2006: Analysis of a textual entailer. *International Conference on Intelligent Text Processing and Computational Linguistics CICLing-06*, Mexico City, Mexico.

Trabasso, Tom, and van den Broek, Paul. 1985. Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612-630.

Vellutino, Frank. 2003. Individual differences as sources of variability in reading comprehension in elementary school children. In *Rethinking reading comprehension*, eds. Anne Sweet and Catherine Snow. New York: Guildford.

Charles Hall,
Department of English
Patterson 467
Memphis, TN 38152
cehall@memphis.edu

Debra S. Lee
CEELI Institute
Havlickovy Sady 58
Prague, the Czech Republic 12000
debraleetn@gmail.com

Philip M. McCarthy
Department of Psychology
Memphis, TN 38152
pmmccrth@memphis.edu

Danielle S. McNamara
Department of Psychology
Memphis, TN 38152
d.mcnamara@mail.psyc.memphis.edu

Gwyneth A. Lewis
Department of Psychology
Memphis, TN 38152
glewis@memphis.edu