

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Language in Law: Using Coh-Metrix to Assess Differences between American and English/Welsh Language Varieties

#### **Permalink**

<https://escholarship.org/uc/item/1vh1t2hr>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

#### **ISSN**

1069-7977

#### **Authors**

Hall, Charles  
Lee, Debra S.  
Lewis, Gwenyth A.  
et al.

#### **Publication Date**

2006

Peer reviewed

# Language in Law: Using Coh-Metrix to Assess Differences between American and English/Welsh Language Varieties

Charles Hall\*, Gwyneth A. Lewis, Philip M. McCarthy, Debra S. Lee\*\*  
Danielle S. McNamara

\* Department of English, Patterson 467  
Memphis. TN 38152 ([cehall@memphis.edu](mailto:cehall@memphis.edu))

Prague, the Czech Republic 12000  
([dlee@leapp.us](mailto:dlee@leapp.us))

\*\* CEELI Institute, Havlickovy Sady 58

Department of Psychology  
Memphis. TN 38152  
{[glewis](mailto:glewis@memphis.edu), [pmmcrth](mailto:pmmcrth@memphis.edu), [dsmcnamr](mailto:dsmcnamr@memphis.edu)} @memphis.edu)

In this study, we add to the limited research on discourse differences between American and British language varieties. We used *Coh-Metrix* (Graesser et al., 2004), an automated tool that can process over 300 indices of cohesion and difficulty, to distinguish a specially constructed written corpus of *law* texts: *American language variety texts* (ALVT) and *English/Welsh language variety text* (EWLVT).

Our corpus, containing 408 *commercial competition* cases (ALVT=200, EWLVT =208), was randomly divided into a *training set* ( $n=200$  texts), and a *test set* ( $n=208$  texts). Using ANOVA performed on the training set, we selected the most significant *Coh-Metrix* predictor indices from each of five distinct categories: *coreferential cohesion*, *casual cohesion*, *local-grammatical cohesion*, *latent semantic analysis*, and *lexical diversity*. We then conducted a discriminant analysis (DA) with language variety as the dependent variable.

## Results and Discussion

The DA derived algorithm when applied to the *test set* correctly categorized 177 of the 208 texts, an average accuracy rate of 85% (ALVT, Precision=.835, Recall=.860; EWLVT, Precision=.867, Recall=.843). As such, this initial

study, offering compelling evidence that significant differences between English language varieties do exist, casts doubt on previous generalizations about British and American writing (Biber, 1987). The study also suggests that language varieties can be computationally distinguished by a tool such as *Coh-Metrix*.

Future research will assess the degree to which expert human raters can distinguish differences between such language varieties. We will also assess whether the differences recorded in this study extend to differences for ALVT and EWLVT when analyzing text material derived from expository and narrative text types.

## Acknowledgements

This research was supported by the Institute for Education Sciences (IES R3056020018-02).

## References

- Biber, D. (1987). A textual comparison of British and American writing. *American Speech*, 62, 99-119.
- Graesser, A., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.