Matching Readers to Texts: LSA and the Goldilocks Principle

Walter Kintsch, M.E. Schreiner, Michael B. W. Wolfe, Bob Rehder
({wkintsch, missy, wolfem, rehder}@psych.colorado.edu)
Department of Psychology & Institute of Cognitive Science
University of Colorado, Boulder Boulder, CO 80309-0345 USA

The Problem and a Potential Solution

When readers set out to learn something from text, they are often faced with the problem of selecting a text from which they will learn as much as possible about the target topic. Of all the possible texts a learner could read, many will be too difficult for the reader to understand, and many will be too simple for the reader to learn anything new (Kintsch, 1994). The goal is to select a text which is 'just right' with respect to what the reader already knows. We refer to this as the Goldilocks Principle.

Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) is a statistical technique which can be used to assess the semantic similarity between two documents. By representing what the reader knows as a document and then comparing that document to other possible documents the reader might read, we can measure the conceptual similarity between what the reader already knows and the texts he or she might read.

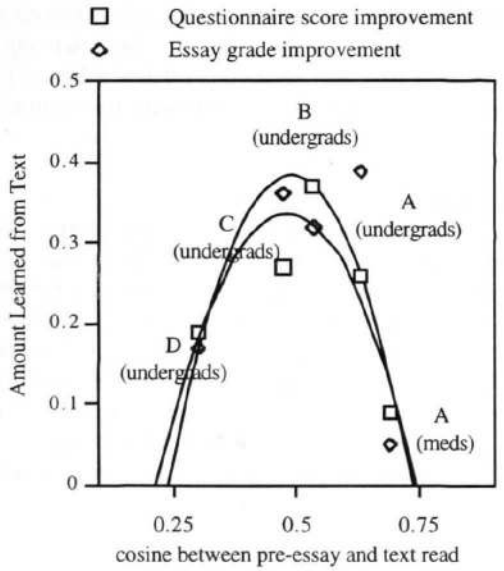

Figure 1. Learning as a function of pre-essay to text similarity for each of the five text conditions.

The Experiment

94 undergraduates and 12 medical students each wrote an essay summarizing their knowledge about the human circulatory system. They then completed a questionnaire about the circulatory system, after which they read one of four texts about the human circulatory system. These four texts varied in difficulty level from elementary school level (Text A) to medical school level (Text D). After reading, the participants did the essay and questionnaire tasks again. (For complete methodological details, see Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch & Landauer, in press). Learning was defined as the proportion improvement on the questionnaire score and on the essay grades.

We expected that learning from a particular text will be non-linearly distributed across knowledge level. Readers who share nothing in common with the target text are expected to learn little: they will find the text too difficult. Readers whose prior knowledge is very similar to the target text are also expected to learn little: they will find the text too easy. It is the readers in the middle who we expect will learn the most: there is enough overlap between their prior knowledge level and the text for them to understand it, and enough non-overlap for them to find new information to learn from it.

The observed relationship between average pre-essay-to-text cosine and learning is given in Figure 1. As expected, learning (as assessed by both the essay grade and the questionnaire score) was non-monotonically related to similarity between initial knowledge and text read.

Conclusions

In this study we found support for the Goldilocks Principle: learning was non-monotonically related to relative prior knowledge. We were able to use LSA to assess conceptual similarity between reader knowledge and text level. The practical implications of this are clear: Wise application of the Goldilocks Principle may make readers more efficient learners. LSA may make application of the Goldilocks Principle pragmatically possible.

Acknowledgments

This research was supported by a grant from the NIMH, MH-15872 to W. Kintsch and a contract from ARPA-CAETI to T. Landauer and W. Kintsch.

References

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49,* 294-303.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104(2),* 211-240.

Wolfe, M. B. W., Schreiner, M.E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W. and Landauer, T. K. (in press). Learning from text: Matching readers and texts by Latent Semantic Analysis. *Discourse Processes.*