

# Automatic Evaluation of Aspects of Document Quality

David F Dufty, Danielle McNamara, Max Louwerse, Ziqiang Cai, Arthur C Graesser

Psychology Department

University of Memphis

Memphis, TN 38152

+1 901 678 2037

d.duffy@mail.psyc.memphis.edu

## ABSTRACT

Coh-Metrix is a web-based application currently in development that automatically evaluates text. It uses two central concepts from discourse processing: text-based cohesion and situation-model based coherence. Cohesion is the degree to which components of the text are linked. Coherence is the representation of the world that the text conveys. Our intention is for Coh-Metrix to eventually map the cohesion of a text to the background knowledge and reading skills of the reader. Coh-Metrix will then be able to give feedback to a writer about which aspects of the text are cohesive and which lack cohesion. This will enable the writer to determine which aspects of the text need to be improved. Applications of Coh-Metrix on document quality as well as other future directions for the development of Coh-Metrix are discussed.

## Categories and Subject Descriptors

J.4 [Computer applications]: application in social and behavioral sciences, psychology

## General Terms

Algorithms, Documentation.

## Keywords

Cohesion; document evaluation, text analysis, writing tools.

## 1. INTRODUCTION

What distinguishes poor writing from good writing and good writing from great writing is the ability of a writer to revise multiple times. In the past, writing ability was taught largely by example, with the hope that the fledgling writer would absorb the appropriate skills by osmosis. Work by theorists such as Flower [3], have done much to make the revision process transparent. There is an extensive literature on text revision and an equally vast array of books outlining techniques that can be used in the revising process [1] [15]. The purpose of this paper is to describe

how advances in the field of discourse processing can inform the revision process, and how computational linguistics can implement tools that can provide insights into the strengths and weaknesses of a document, in particular the cohesion of the document.

According to Flower [3], early versions of a document are writer-based. The goal of the writer is to move the text from being writer-based to being reader-based. Writer-based text is a necessary stage of writing, as it is the product of the writer's initial attempts to put ideas into words and form some kind of global framework. However, such text is often difficult to read. Writer-based text tends to be erratic, jumping from idea to idea. It is often unstructured, poorly expressed, and has extensive use of 'private language' - words and phrases with meaning to the writer but which are either ambiguous or unclear to the reader.

After the creation of a writer-based first draft, the next step is a critical evaluation of the document, either by the writer or another person. The following advice from Axelrod and Cooper [1] (p. 212) is directed at an outside reviewer, but is also intended for writers to use in objective evaluation of their own work:

Evaluate the Organization. Look at the way the essay is organized by making a scratch outline. Does the information seem to be logically divided? If not, suggest a better way to divide it. Also consider the order or sequence of information. Can you suggest a better way of sequencing it?

This suggestion comes from a much longer list of techniques for critical evaluation of an early draft. Such advice, while valuable, requires a high level of expertise in critical evaluation as well as considerable time and effort. The revision process can be aided by computational tools for evaluating text.

### 1.1 Measures of text quality

Until recently, evaluation of the readability of a text (reader-based) primarily consisted of programs that assign a single score to a text, providing no specific guidance to the writer about how to improve their text. Quantitative measures of text quality initially focused on simple surface characteristics of the text, such as word length and sentence length. Perhaps the most influential of these are the Flesch Reading Ease Score and the Flesch-Kincaid Grade Level. Both of these measures combine number of syllables per word and average sentence length to produce a readability measure. The Flesch Reading Ease produces a score between 0 and 100 with higher numbers indicating easier texts. The Flesch-Kincaid Grade Level assigns a number between 1 and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGDOC '04, October 10–13, 2004, Memphis, Tennessee, USA.

Copyright 2004 ACM 1-58113-809-1/04/0010...\$5.00.

12 that is intended to be an approximation to the appropriate grade level of readers of the text in question. Other measures have incorporated features such as word frequency [17].

More recently, systems have been developed that make use of more complex algorithms and greater computing power, such as the Intelligent Essay Assessor [4] [16] and e-rater [2].

Intelligent Essay Assessor makes use of Latent Semantic Analysis, or LSA [9], a technique for comparing the similarity of one text with another. An LSA comparison between two texts gives a number between 0 and 1, which indicates the similarity of the two texts, where 1 means the texts are identical and 0 means the texts have no similarity. The Intelligent Essay Assessor evaluates text by comparing a student's essay against an ideal essay. Furthermore, the system can give more information than earlier readability scores that just gave a single measure of quality. If particular sub-topics are identified in the ideal essay prior to analysis, the Intelligent Essay Assessor can compare the student's essay to each of the sub-topics, and give feedback to the student about subtopics that they have missed or covered inadequately [5].

E-rater is an automated essay rating system with a focus on evaluation of completed essays rather than feedback during the writing process. E-rater analyzes an essay on three levels: rhetorical structure theory [11], syntactic structure, and topical analysis. Rhetorical structure is evaluated with algorithms that detect parallelism and contrast. Syntactic structure is measured by evaluating the variety of syntactic constructions used, with more variety producing higher scores. Topical analysis uses word overlap to determine the degree to which the essay stays on topic.

## 2. COH-METRIX

Coh-Metrix [6] [7] is a web-based text assessment application currently in development. Coh-Metrix uses two central concepts from discourse processing: text-based cohesion and mental-model based coherence. Recent evidence from psychology and linguistics shows that an important factor in text comprehension is cohesion [6] [10] [14]. Cohesion is the degree to which components of the text are linked. Cohesion is thus distinct from coherence, which is the extent to which a reader develops a unified situation model of the text. Cohesion may be interpreted as the inverse of the number of inferences required by a reader to link the elements of a text into a coherent situation model.

The marriage of cohesion and coherence in a computational tool provides algorithms that give specific information about the strengths and weaknesses of a document.

According to current theories of discourse, a text has at least three levels of structure: the surface code, the text-base, and the situation model [8]. The surface code consists of the literal words used in the text, the text-base consists of the propositions that the surface code describes, and the situation model is the representation of the world that the text is intended to convey. The effort involved in constructing a situation model from the text is critical in distinguishing easy texts from difficult ones and good quality writing from poor. Situation models are considered to be multidimensional. Readers represent the situation in the text on at least the dimensions of time, space, causality, intentionality, and agency [18]. The background knowledge and reading skills of the reader play an important role in developing a coherent situation model [13] [14].

### 2.1 Dimensions of cohesion

An assumption behind Coh-Metrix is that cohesion can be measured separately for each dimension of the situation model. That is, a text will have separate cohesion levels for causal, intentional, temporal, referential, spatial, and structural cohesion. Coh-Metrix estimates the cohesion of text on each of these dimensions. The full range of measures that Coh-Metrix calculates are covered in detail elsewhere [7]. Instead, a short passage from a current school text will be used as an example to demonstrate how Coh-Metrix may facilitate the revision process with a focus on the dimension of causal cohesion.

The following segment was taken from a 4th grade science text. The segment comes from a larger passage describing the effects of heat on matter.

Adding or taking away heat can change matter. Matter is something that takes up space. Matter can change from one state, or form, to another.

An ice cube is solid water. Solid is one state of matter. Heat can melt an ice cube. The ice cube changes into liquid water. Liquid is another state of matter. When heat is taken away, the water can change back. Liquid water turns into solid water.

Heat can make liquids boil. Water boils when it is heated. When the water boils, it turns into a gas. This gas is called water vapor. Solid, liquid and gas are three states of matter.

The Flesch-Kincaid Grade Level for the text, calculated from the average number of syllables per word and the average sentence length, is 3.2. This low grade level rating does not seem consistent with the difficulty of the text or its readability. The passage is disjointed and seems to present a cavalcade of miscellaneous facts. The reason for the disparity between the calculated readability score and the confusing impression the passage creates is the reliance on sentence length and word length to measure the readability of a text. In the case of the text here, the sentences are short and choppy, and thus produce a low Flesch-Kincaid Grade Level score. If that score alone were used, the text would be rated easy to read, and would be in need of no further modification. The cohesion scores for the passage, however, tell a different story.

The passage is replete with causal information. The central purpose of the passage is to describe how the presence or absence of heat can cause events to occur such as ice melting or water freezing. Coh-Metrix captures this dense causal information by detecting causal verbs such as melt and change. However, the causal information is poorly structured, resulting in a low causal cohesion score. None of the causal relationships in the text are explicit to the reader but instead must be inferred. This places demands on the working memory of the reader when the reader is trying to build a situation model of the text. For example, the causal action of heat on ice and water is not explicitly stated, but must be constructed from information in successive sentences. The high amount of causal information coupled with low causal cohesion indicates that the text needs to be more cohesive and explicit in describing causal events.

If the writer of this passage had received feedback on the causal cohesion of the passage, they may have improved it by re-organizing the causal information and making the causal

connections in the text explicit. A variation of the text with improved causal cohesion is the following:

Adding heat or taking away heat can change matter. Matter is something that takes up space. Matter can change from one state to another state, or from one form to another form. Three states of matter are solid, liquid and gas. For example, an ice cube is solid water. Heat can melt an ice cube, causing the ice cube to change into liquid water. When heat is taken away, the liquid water

Other dimensions of cohesion for the passage are either not low or not important. For example, intentional cohesion is the extent to which the goals and intentions of agents described in the text are explicit. Intentional cohesion as measured by Coh-Matrix is effectively zero in the passage. However, as there are no animate nouns, that there is no intentional information, and therefore the level of intentional cohesion is not important to the overall structure of the text.

### 3. FUTURE WORK IN DEVELOPMENT OF COH-MATRIX

It is important to note that documents with the highest levels of cohesion are not necessarily the best. High cohesion can sometimes impair comprehension for readers with high levels of prior knowledge of the topic [13] [14]. Therefore, the optimum level of cohesion is dependent on the knowledge base of the audience. We are currently undertaking empirical studies and corpus linguistic research to determine the mapping between cohesion and coherence.

A second text computation tool is also being developed, and is referred to as Coh-GIT (Cohesion Gap Identification Tool). While Coh-Matrix computes overall cohesion, Coh-GIT will identify specific cohesion gaps in the text. These two tools will bring a previously unseen level of sophistication to automated text evaluation and provide assistance to the writing process.

Computational techniques such as those currently being implemented in Coh-Matrix will provide a writers with new tools for improving text as well as providing educators with automated measures of giving feedback on specific strengths and weaknesses of students' writing.

### 4. ACKNOWLEDGMENTS

The research was supported by a grant from the Institute of Education Sciences (IES R3056020018-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

### 5. REFERENCES

- [1] Axelrod, R. B., & Cooper, C. R. (2001). *The St Martins Guide to Writing*. Bedford/St Martins: Boston, MA.
- [2] Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., & Harris, M. D. (1998). Automated Scoring Using A Hybrid Feature Identification Technique. *Proceedings of the Annual Meeting of the Association of Computational Linguistics, August 1998*. Montreal, Canada.
- [3] Flower, L. (1993). *Problem-Solving Strategies for Writers. (4th Ed)*. Fort Worth: Harcourt Brace Jovanovich
- [4] Foltz, P., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25*, 285–307.
- [5] Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The Intelligent Essay Assessor: Applications to Educational Technology. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 1(2)*
- [6] 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 (pp. 82-98)*. New York: Guilford Press.
- [7] Graesser, A.C., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Matrix: Coh-Matrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*. In press.
- [8] Kintsch, W. (1998) *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- [9] Landauer, T., and S. Dumais. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.
- [10] Louwerse, M.M. (2002). An analytic and cognitive parameterization of coherence relations. *Cognitive Linguistics, 12*, 291–315.
- [11] Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text, 8(3)*, 243-281.
- [12] McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). *Coh-Matrix: Automated cohesion and coherence scores to predict readability and facilitate comprehension*. Unpublished technical report: University of Memphis.
- [13] McNamara, D. S., & Kintsch, W. (1996). Learning from Text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-287.
- [14] McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.
- [15] Reid, S. (1997). *The Prentice-Hall guide for college writers: a customized edition*. Pearson Custom Publishing: Needham Heights, MA.
- [16] Schreiner, M., Rehder, B., Landauer, T., & Laham, D. (1997). How latent semantic analysis (LSA) represents essay semantic content: Technical issues and analysis. In M. Shafto and P. Langley (Eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society (pp 1041)*. Mahwah, NJ: Erlbaum.
- [17] Stenner, A. J. (1998). Measuring reading comprehension with the lexile framework. *4th North American Conference on Adolescent/Adult Literacy*: Washington, D.C.
- [18] Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.