IIS: A Marriage of Computational Linguistics, Psychology, and Educational Technologies

Danielle S. McNamara
University of Memphis
d.mcnamara@mail.psyc.memphis.edu

Abstract

There are numerous projects at the University of Memphis Institute for Intelligent Systems (IIS) that involve a *marriage* of natural language processing, computational linguistics, psychology, and educational technologies. We use natural language and computational linguistic technologies to accomplish a variety of goals related to psychology and educational practice, including text analysis and the development of interactive tutoring systems. This paper describes these technologies and some of our computational methods.

Introduction

Conferences such as FLAIRS excel in interdisciplinary research. Likewise, that is exactly what defines the Institute for Intelligent Systems (IIS) at the University of Memphis. One aspect of our mission involves using multiple, interdisciplinary approaches to answer questions related to cognitive science and educational practice. These methods include behavioral experiments as well as methods such as computational modeling and computational linguistics. These approaches and methods form an interdisciplinary collaborative marriage.

The projects at the IIS can be divided into two mutually supportive camps. One includes tutoring technologies. These technologies are being developed to support learning, but also to enhance our theoretical understanding of various cognitive processes such as knowledge acquisition and reading comprehension. The other camp of projects at the IIS centers on discourse technologies. These projects vary from automated tools that provide information on hundreds of features of text to projects investigating multimodal communication between humans.

The latter camp supports the former by providing a means to hold natural language dialogues with the students who interact with the tutoring systems. All of the tutoring systems at the IIS include interactive dialogue between the user and pedagogical agents that guide the learning process. These tutoring technologies rest on the principles that active processing is critical to learning and that interactive dialogue scaffolds that level of processing.

Tutoring Technologies at IIS

AutoTutor

The AutoTutor project, led by Art Graesser, has been investigating tutoring, tutoring technologies, and natural language processing for over a decade (e.g., Graesser et al., 2005). AutoTutor is a *talking head* or pedagogical agent that guides students in learning about topics such as physics, computer literacy, and critical thinking. It simulates a human tutor by holding a conversation with the learner and presenting series of questions or problems for the student to answer. For example, one question in conceptual physics is *"When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation?"* Although the ideal answer to this question is approximately a paragraph in length, the initial answer by a typical student generally varies between a single word and a couple of sentences. AutoTutor goes through a series of *questions*, *prompts*, and *hints* to guide the learner to the correct and complete answer to the question. This tutorial dialogue typically lasts 50-200 turns (i.e., the learner expresses something, then AutoTutor, then the learner, and so on), which is generally comparable to what can be expected during an interaction with a human tutor.

AutoTutor accomplishes this seemingly natural dialogue with the student by segmenting the student's turns into speech act units. It then assigns these units into one of 20 categories, such as *Assertion*, *Short Answer*, *Meta-communication*, *Verification Question*, and *Comparison Question*. Interpreting and responding to the student's speech acts is achieved by implementing modules developed in the field of computational linguistics. The language interpretation component implements a variety of natural language understanding technologies, including simple word matching, part-of-speech tagging, frozen expression matching, regular expression recognizers, syntactic parsers, and Latent Semantic Analysis (LSA; Landauer & Dumais, 1997; Landauer et al., 2007).

Based on the interpretation of what the student has input into the system, AutoTutor must then respond to the student. The language generation component has a goal-based dialogue move planner and natural language generation templates. Each turn of AutoTutor requires the generation of one or more dialogue moves that adaptively

respond to what the student just expressed and that advance the conversation in a constructive fashion. Some dialogue moves respond to the student's preceding turn, such as short feedback (positive, neutral, negative), answers to student questions, and corrections of student misconceptions. Other dialogue moves push the dialogue forward in an attempt to cover the expected answer to the main question. These forward-directed dialogue moves include Pumps (e.g., *Tell me more, What else?*), Hints, Prompts for specific words or phrases, and Assertions.

AutoTutor's primary method of scaffolding the students' answers is *Expectation and misconception tailored dialogue* (EMT dialogue). This includes a list of anticipated good answers (called *expectations*) and a list of *misconceptions* associated with each question. One goal of AutoTutor is to guide the student in covering the list of expectations. This goal is accomplished with forward-directed dialogue moves until the student or tutor covers each expectation on the list. A second goal is to correct the student's misconceptions, generally by correcting the errors as soon as they occur. A third goal is to adaptively respond to the student by giving feedback on the quality of student contributions and by answering the student's questions. A fourth goal is to manage the dialogue in a fashion that appears coherent and accommodates unusual speech acts by learners. As the learner expresses information over many turns, the list of expectations is eventually covered and the main question is scored as answered. Then, AutoTutor moves on to the next question.

iSTART

iSTART (Interactive Strategy Training for Active Reading and Thinking; McNamara, Levinstein, & Boonthum, 2004) is an interactive tutoring system designed to help students learn reading comprehension strategies that support deeper understanding of challenging text (McNamara et al., 2004). iSTART's design is based on a successful classroom intervention called Self-Explanation Reading Training (SERT; McNamara, 2004). SERT combines the power of self-explanation in facilitating deep learning (Chi et al., 1994) with content-sensitive, interactive strategy training (Bereiter & Bird, 1985).

iSTART teaches readers to self-explain using five reading strategies: *monitoring comprehension* (i.e., recognizing comprehension failures and the need for remedial strategies), *paraphrasing* explicit text, making *bridging inferences* between the current sentence and prior text, making *predictions* about the subsequent text, and *elaborating* the text with links to what the reader already knows. The system uses full bodied, animated agents to scaffold the learning of these comprehension strategies (McNamara et al., 2007; McNamara et al., 2004).

The animated agents in iSTART provide three phases of training. The *Introduction Module* provides instruction on self-explanation and reading strategies. There is a trio of animated agents (an instructor and two students) that cooperate with each other, provide information, pose questions, and provide explanations of the reading strategies. After the presentation of each strategy, the trainees complete brief multiple-choice quizzes to assess their learning. In the *Demonstration Module*, two Microsoft Agent characters (Merlin and Genie) demonstrate the use of self-explanation in the context of a science passage and the trainee identifies the strategies being used. The final phase is *Practice*, where Merlin coaches and provides feedback to the trainee while the trainee practices self-explanation using the repertoire of reading strategies. For each sentence in a text, Merlin reads the sentence and asks the trainee to self-explain it by typing a self-explanation. Merlin gives feedback and asks the trainee to modify unsatisfactory self-explanations.

Like AutoTutor, iSTART interprets the trainees' contributions using computational linguistic algorithms. The systems differ in their underlying needs vis-à-vis the algorithms that support that process. AutoTutor must assess the *accuracy* of the content of the students' input. In contrast, iSTART must assess the nature of *strategies* that the student uses to understand the text being read.

Feedback on the content and quality of the self-explanation is a critical component of practice. During the practice phase, the agents' interactions with the trainee are moderated by the quality of the explanation: more positive feedback is given for longer, more relevant explanations, whereas increased interactions and support are provided for shorter, less relevant explanations. The computational challenge is for the system to provide appropriate feedback within seconds to the student concerning the quality of the self-explanations.

McNamara and colleagues (2007) described the evolution of the iSTART algorithms. The first version of iSTART relied solely on word matching and benchmark words to guide feedback. The current system uses a variety of sources of information, including the text being read and self explained by the reader as well as analyses using LSA. First, the response is screened for metacognitive expressions (e.g., *I don't understand what they are saying*). The expressions are located using pattern recognition based on regular expressions and a collection of patterns derived from explanations collected during experiments. The types of metacognitive expressions and their lengths are used to direct the system's response to the student.

Second, the remainder of the explanation is analyzed using word-matching and LSA. The algorithms computed for the word-based evaluation of the self-explanation use the length of student's self-explanation relative to that of the target sentence and the number of content words that match content words from the target sentence and the rest of the text. These algorithms determine if the explanation is irrelevant to the sentence, too short to be a reasonable explanation, or too similar to the target sentence.

LSA is a method for representing word meanings from a large corpus of text (e.g., Landauer et al., 2007). It uses a

statistical analysis to infer semantic similarity among words. The current evaluation system in iSTART calculates the cosines between the student's explanation and collections of words representing various parts of the passage being explained: the target sentence, title, and prior content.

The results from the word-based and LSA-based analyses are integrated with the metacognitive screening to produce feedback in one of the following six categories: 1) response to the metacognitive content, 2) feedback that the explanation appears irrelevant to the text, 3) feedback that the explanation is too short compared to the content of the sentence, 4) feedback that the explanation is too similar to the original, 5) feedback that makes a suggestion for the following sentence, or 6) feedback that gives an appropriate level of praise.

The success of the evaluation algorithms has been evaluated by computing linear equations based on discriminate analysis of one data set and assessing its ability to predict ratings for a variety of data sets (McNamara et al., 2004; McNamara et al., 2007, Millis et al., 2003). We have compared the results from discriminate functions using word-based variables alone, LSA cosines alone, and the combination of LSA and the word-based variables. Signal detection (d prime) is used to assess the goodness of prediction for each of these algorithms. The general trend in the performance of these algorithms has been that LSA/word-based combination performs better than word-based algorithms, which in turn captures more than LSA-based algorithms alone. The d-primes in these studies have ranged from 1.0 to 3.0. Thus, across a number of evaluations, the iSTART algorithms have corresponded well to human ratings.

MetaTutor

MetaTutor is a new project led by Roger Azevedo. It focuses on teaching students self-regulated learning strategies in the context of learning about science topics such as the human circulatory system. The underlying assumption of MetaTutor is that students should regulate key cognitive, metacognitive, motivational, social, and affective processes in order to learn about complex science topics. The design of MetaTutor is based on extensive research by Azevedo and colleagues showing that providing adaptive human scaffolding that addresses both the content of the domain and the processes of self-regulated learning enhances students' learning about a challenging science topic hypermedia (e.g., Azevedo & Cromley, 2004, Azevedo et al., 2005). Azevedo's team has investigated dozens of self-regulatory processes related to *planning* (e.g., generating sub-goals, activating knowledge), *monitoring activities*, effective and ineffective *learning strategies*, and methods of *handling task difficulties and demands* (e.g., time and effort planning).

The purpose of the MetaTutor project is to examine the effectiveness of animated pedagogical agents as external regulating agents used to detect, trace, model, and foster students' self-regulatory processes during learning about complex science topics. MetaTutor will provide instruction to learn SRL strategies and provides scaffolding to use those strategies while learning about the circulatory system.

MetaTutor is in its infancy, thus the algorithms to guide feedback to the student are not yet developed and tested. The challenge will be to provide feedback both on the accuracy of the content as well as the appropriateness of the strategies used by the student. Thus, MetaTutor will be somewhat of an amalgamation of AutoTutor and iSTART.

IIS Discourse Technologies

Another side of IIS involves the analysis of text and discourse. These projects include an array of automated tools, algorithms, and experimental investigations, but all of them are directed at better understanding how to improve text, discourse, and communication.

QUAID

QUAID (Question Understanding Aid) is a project led by Art Graesser that preceded and largely laid the groundwork necessary to build many of the tools described in this paper. It is a software tool designed to assess the quality of questions. Its purpose is to help survey methodologists, social scientists, and designers of questionnaires to improve the wording, syntax, and semantics of questions. QUAID (available on the internet) identifies potential problems that might impede question comprehension, such as unfamiliar terms, vague or ambiguous noun phrases, and complex syntax (Graesser et al., 2006).

Coh-Metrix

Perhaps the largest project currently being conducted at IIS is Coh-Metrix (cohmetrix.memphis.edu). Our goals include investigating the effects of cohesion on comprehension and developing a computational tool that provides hundreds of indices of cohesion, difficulty and readability (Graesser et al. 2004; McNamara et al., 2006). Coh-Metrix measures texts on a broad profile of language, cohesion, and conceptual characteristics. The system integrates lexicons, pattern classifiers, part-of-speech taggers (Brill, 1995), syntactic parsers (Charniak 2000), shallow semantic interpreters, LSA, and other components that have been developed in the field of computational linguistics. Coh-Metrix taps into a large number of commonly used lexicons including Celex (Baayen, Piepenbrock, & Van Rijn, 1993), WordNet (Miller et al, 1990), the MRC Psycholinguistic Database (Coltheart, 1981), and lexicons that estimate the frequency of word usage in the English language. These lexicons allow us to measure each word on number of features, including syllables, abstractness, imagery, ambiguity, frequency of usage, age of acquisition, number of senses (meanings), and dozens of other

dimensions. Coh-Metrix incorporates a syntactic parser that assigns parts of speech and syntactic tree structures to sentences and measures them on syntactic complexity. Another module uses LSA to measure the conceptual similarity between sentences, paragraphs, and texts on the basis of world knowledge. Coh-Metrix also provides researchers with a range of traditional indices such as average word length, average sentence length, and readability formulas such as the *Flesh Reading Ease*.

While there are an abundance of indices provided by Coh-Metrix, the primary purpose of the tool is to provide indices of cohesion (McNamara et al., 2006), which is key to comprehension (e.g., McNamara, 2001). These include indices of referential cohesion, causal, intentional, spatial, temporal, and structural cohesion. For example, Coh-Metrix provides a variety of measures of referential cohesion, including noun overlap, argument overlap, stem overlap, and LSA-based semantic overlap. These indices indicate the degree of word and conceptual overlap between sentences, paragraphs, and within the entire text.

Coh-Metrix has been used for a variety of purposes. The indices have successfully distinguished between types, domains, genres, and modes of texts. For example, we have developed and tested measures of genre identification (Dempsey, McCarthy, & McNamara, 2007; Lightman, McCarthy et al., 2007; McCarthy, Graesser, & McNamara 2006), temporal cohesion (Duran et al., 2006), topic sentencehood (McCarthy, Renner, et al., 2007), and mode of the text (i.e., spoken vs. written: Louwerse et al., 2004). Crossley et al. (2007) used Coh-Metrix to identify differences between authentic and simplified versions of texts designed for English language learners. McCarthy, Briner, et al. (2007) used Coh-Metrix to differentiate sections of science texts. Also, Coh-Metrix distinguishes between authors (McCarthy, Lewis, et al., 2006) and further identifies the nationality (Hall et al., 2007; McCarthy, Lehenbauer, et al., 2007), gender (Bell, McCarthy, & McNamara, 2006), and emotional state (Lightman, McCarthy, & McNamara, 2007) of the author.

A number of projects and tools have already emerged from the Coh-Metrix project, and a number of projects are envisioned. For example, we have developed a prototype writing aid tool, called *WRAID* (Writing and Reading Aid), designed to identify problems with the wording, syntax, and semantics of a text. We also envision a tool called *CoGIT* (Cohesion Gap Identification Tool) to identify cohesion gaps in text. The ultimate goal of Coh-Metrix is to match reader to text, by identifying the features of text that are most beneficial for a particular reader.

Entailer

The *Entailer* project (Rus et al., 2005), led by Vasile Rus, involves the development of algorithms to evaluate the degree to which one text is entailed by another text. For example, *John drove to the store to buy supplies*, entails, *John went to the store*. The Entailer emerges from the testing ground of the *recognizing textual entailment* corpus (RTE; Dagan, Glickman, & Magnini, 2004-2005). However, in our world of tutoring systems, the Entailer potentially offers alternative methods for distinguishing subtle differences between tutees' responses. A major challenge for tutoring systems that incorporate natural language interaction is to accurately evaluate users' contributions and to produce accurate feedback (McCarthy, Rus, et al., 2007). Entailer further addresses that goal.

Entailer encompasses lexico-syntactic information, negation handling (*explicit* and *implicit)*, synonymy, and antonymy. It functions by having each pair of text fragments (assigned as text [T] and hypothesis [H]) mapped into two graphs, one for T and one for H, with nodes representing main concepts and links indicating syntactic dependencies among concepts as encoded in T and H, respectively. An entailment score, entail(T,H), is then computed quantifying the degree to which the T-graph subsumes the H-graph. The score is the weighted sum of one lexical and one syntactic component. The lexical component expresses the degree of subsumption between H and T at word level, (i.e., vertex-level) while the syntactic component works the same way at syntactic level (i.e. edge-level).

For the purposes of natural language assessment in ITS, *Entailer* offers a number of advantages over current text relatedness measures such as LSA and overlap indices. First, because lexical/word information acts only as a component of the overall formula, *Entailer* is less susceptible to the problem of text length. Additionally, as *Entailer* addresses both syntactic relations and negation, the tendency for higher relatedness results over lengthier texts is reduced. Second, *Entailer* evaluates the relation between text and hypothesis non-symmetrically, such that entail (T,H) does not entail (H,T). As such, the evaluation of a response to a stimulus or source text will be different from the evaluation of the stimulus to the response. Third, *Entailer* handles negations; so it offers the opportunity of providing more accurate feedback. Entailer results thus far have been promising and better than state-of-the-art solutions that use the same array of resources (e.g., Rus et al., 2005).

Intelligent MapTask Agent

The projects discussed thus far have centered primarily on language. Human communication, however, involves a rich combination of communicative channels, including speech features, facial expressions, eye gaze, and hand gestures. Max Louwerse leads a project that investigates this wider scope of multimodal communication. Louwerse and his team are investigating the interaction between discourse, speech, eye gaze, facial expression, gesture, and map drawing (Louwerse et al., 2006). The project aims to determine how these modalities are aligned, whether and when these modalities are observed, and whether the correct use of these channels actually aids comprehension.

Due to the inherent complexity of multimodal communication, controlling for genre, topic, and goals during unscripted dialogue is crucial. With these concerns in mind, they used the Map Task scenario, a restricted-domain, route-communication task. In the Map Task scenario, it is possible for experimenters to know exactly what each participant knows at any given time. In this scenario, the Instruction Giver coaches the Instruction Follower through a route on the map. By way of instruction, participants are told that they and their interlocutors have maps of the same location, but drawn by different explorers, and so potentially different in detail. They are not told where or how the maps differ. Data from each conversation consists of recordings of participants' facial expressions, gestures, speech, eye gaze patterns (for the Giver), and map drawings (for the Follower).

Based on this research, an initial embodied conversational system has been developed that uses speech recognition to interactively guide the dialogue partner through the route on a map. In a similar vein, one current goal within the AutoTutor project is to incorporate more communication channels into the feedback system. They are particularly focusing on the role of emotion, and how its assessment may enhance feedback. Understanding our multiple channels of communication will ultimately enhance our understanding how communicative resources are used in face-to-face conversations and how animated conversational agents can become more human-like.

Conclusion

This overview provides a glimpse of the various interdisciplinary research projects at the IIS. These projects center on natural language processing techniques, employing a wide variety of computational approaches to address a variety of problems and questions related to psychology and education. And, vice versa, we use a variety of approaches germane to psychology to solve problems related to natural language processing. We hope that this marriage of interdisciplinary approaches and methodologies provides the research community with a plethora of tools, research findings, and theories.

Acknowledgements

References

Azevedo, R., and Cromley, J.G. 2004. Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523-535.

Azevedo, R., Cromley, J.G., Winters, F.I., Moos, D.C., and Greene, J.A. 2005. Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science, 33*, 381-412.

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (Eds.) 1993. The CELEX Lexical Database [CD-ROM]. University of Pennsylvania, Philadelphia: Linguistic Data Consortium.

Bell, C.M., McCarthy, P.M., and McNamara, D.S., 2006. Variations in language use across gender: Biological versus sociological theories. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1009-1013). Mahwah, NJ: Erlbaum.

Bereiter C. and Bird, M. 1985. Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2,* 131-156.

Brill, E 1995. Transformation-Based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*.

Charniak, E. 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.

Chi, M. T. H., deLeeuw, N., Chiu, M., and LaVancher, C. 1994. Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

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

Crossley, S.A., Louwerse, M.M., McCarthy, P.M. and McNamara, D.S. 2007. A linguistic analysis of simplified and authentic texts. Modern Language Journal, 91.

Dagan, I., Glickman, O., and Magnini, B. 2004-2005. The PASCAL Recognizing textual entailment challenge. *Proceedings of the Recognizing Textual Entailment Challenge Workshop*.

Dempsey, K.B., McCarthy, P.M., and McNamara, D.S. 2007. *Using phrasal verbs as an index to distinguish text genres*. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)*.

Duran, N., McCarthy, P.M., Graesser, A.C., and McNamara, D.S., 2006. Using Coh-Metrix temporal indices to predict psychological measures of time. Proceedings of the 28th Annual Conference of the Cognitive Science Society (pp. 190-195). Mahwah, NJ: Erlbaum.

Graesser, A.C., Chipman, P., Haynes, B.C., and Olney, A. 2005. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education* 48, 612-618.

Graesser, A.C., Cai, Z., Louwerse, M., and Daniel, F. 2006. *Question Understanding Aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. Public Opinion Quarterly, 70,* 1-20.

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

Hall, C, McCarthy, P. M., Lewis, G.A., Lee, D.S., and McNamara, D. S. 2007. A Coh-Metrix Assessment of American and English/Welsh Legal English. *Coyote Papers: Psycholinguistic and Computational Perspectives. University of Arizona Working Papers in Linguistics, 15.*

Landauer, T.K., and 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*, 211-240.

Landauer, T., McNamara, D.S., Dennis, S, and Kintsch, W. (Eds.) 2007. LSA: *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Lightman, E.J., McCarthy, P.M., Dufty, D.F., and McNamara, D.S. 2007. *The structural organization of high school educational texts. Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS).*

Lightman, E.J., McCarthy, P.M., and McNamara, D.S. 2007. Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. Submitted to Cognitive Science, 2007.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., and Graesser, A. C. 2004. Variation in language and cohesion across written and spoken registers. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 843-848). Mahwah, NJ: Erlbaum.

Louwerse, M. M, Jeuniaux, P., Hoque, M. E., Wu, J., and Lewis, G. 2006. Multimodal communication in computer-mediated map task scenarios. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1717-1722)*. Mahwah, NJ: Erlbaum.

McCarthy, P.M., Briner, S.W., Rus, V., and McNamara, D.S. 2007. Textual Signatures: Identifying texttypes 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., Lehenbauer, B.M., Hall, C., Duran, N.D., Fujiwara, Y., and McNamara, D.S. 2007. A Coh-Metrix analysis of discourse variation in the texts of Japanese, American, and British Scientists. *Foreign Languages for Specific Purposes, 6.*

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

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

McCarthy, P.M., Renner, A.M., Duncan, M.G., Duran, N.D., Lightman, E.J., and McNamara. D.S., 2007. A comparison of two models to computationally identify topic sentencehood. *Submitted to Behavior Research and Methods.*

McCarthy, P.M., Rus, V., Crossley, S.A., Bigham, S.C., Graesser, A.C., and McNamara, D.S. 2007. Assessing the *Entailer* with a corpus of natural language from an intelligent tutoring system. *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS).*

McNamara, D.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 D. S. 2004. SERT: Self-explanation reading training. *Discourse Processes*, 38, -30.

McNamara, D. S., Boonthum, C., Levinstein, I. B., and Millis, K. 2007. Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

McNamara, D. S., Levinstein, I. B., and Boonthum, C. 2004. iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instrument, & Computers*, 36, 222-233.

McNamara, D.S., Ozuru, Y., Graesser, A.C., and Louwerse, M. 2006. Validating Coh-Metrix. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 573-578). Mahwah, NJ: Erlbaum.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. 1990. *Five papers on WordNet* (Technical Report). Princeton University, Cognitive Science Laboratory.

Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., and McNamara, D. S. 2007. Assessing and improving comprehension with Latent Semantic Analysis. In Landauer, T., D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Rus, V., Graesser, A.C., McCarthy, P. M., and Lin, K. 2005. A lexico-syntactic approach to textual entailment. *Proceedings of International Conference on Tools with Artificial Intelligence.*