AutoTutor's Log Files and Categories of Language and Discourse

**Johanna Marineau[1], Andrew Olney[1], Max Louwerse[1],
Natalie Person[2], Brent Olde[1], Suresh Susarla[1], Patrick Chipman[1], Art Graesser[1], and the Tutoring Research Group**

**[1]Department of Psychology, 202 Psychology Building,
The University of Memphis, Memphis, TN 38152 USA
jmarinea@memphis.edu**

**[2]Department of Psychology, 2000 North Parkway,
Rhodes College, Memphis, TN 38112 USA**

Abstract: AutoTutor is a computer tutor that interacts with students in a multi-turn tutorial dialog with natural language. AutoTutor has been developed for computer literacy and conceptual physics, showing learning gains of .6 standard deviation units. This paper documents AutoTutor's log files. The log files record pertinent information about each turn that the student contributes, including syntactic parsing, speech act segmentation and classification, coverage of the expected content, the quality of the student contribution, and other dimensions of language and discourse. The log file also records relevant information about the selection of tutor's dialogue moves through the Dialogue Advance Network (DAN).

Keywords: log files, intelligent tutoring systems, annotation, mark-up initiatives, dialog, AutoTutor

Introduction

AutoTutor is a computer tutor that holds conversations with students in natural language (Graesser, Person, Harter, & TRG, 2001; Graesser, VanLehn, Rose, Jordan, & Harter, 2001; Graesser, P. Wiemer-Hastings, K. Wiemer-Hastings, Kreuz, & TRG, 1999). In each tutoring session, AutoTutor presents a number of questions (or problems) that require deep reasoning, explanations, and approximately a paragraph of information in the answer (or solution). A multi-turn exchange between the tutor and student is needed for a full answer to evolve. AutoTutor facilitates the evolution of an answer by generating a number of different categories of dialog acts, such as pumps ("Tell me more."), prompts for specific information, hints, assertions, summaries, corrections, and feedback about the student's contributions (positive, neutral, negative). These dialog acts are sensitive to the quality and nature of student contributions. There also is an animated conversational agent (i.e., talking head) that presents the tutor's contributions.

The argument has been made that that there is something about natural language and discourse that appears to be effective in promoting learning gains in human tutoring (Corbett, Anderson, Graesser, Koedinger, and VanLehn, 1999; Graesser, Person, & Magliano, 1995). This has motivated a number of researchers to incorporate natural language dialog facilities in their tutoring systems (Aleven, Popescu, & Koedinger, 2001; Freeman, 2000; Moore, 1995; Rose, Jordan, Ringenberg, Siler, VanLehn, & Weinstein, 2001). Indeed, we have found that AutoTutor produces a learning gain of .6 standard deviation unit compared to a reread control in our evaluations of AutoTutor in the domain of introductory computer literacy.

It is beyond the scope of this paper to describe the mechanisms that process student contributions, generate dialog moves, and manage the dialog. AutoTutor uses Latent Semantic Analysis (Landauer and Dumais, 1997; Landauer, Foltz, and Laham, 1998) to represent world knowledge and evaluate the conceptual quality of student contributions in comparison with expected answers to deep-reasoning questions (Graesser, K. Wiemer-Hastings, P. Wiemer-

Hastings, Harter, Person, and the TRG, 2000; Franschetti, Karnavat, Marineau, Callie, Olde, Terry, and Graesser, 2001). A Dialogue Advancer Network (DAN) serves as a dialog management facility to handle mixed-initiative dialogue (Person, Graesser, Kreuz, Pomeroy, and the TRG, 2001). There is curriculum script that stores concepts, the sets of deep-reasoning questions, expected answers to questions, families of hints and problems for each expected answer, sets of misconceptions, corrections of errors, and other content about the subject matter. The information in the student's turn is compared to the expected answers and misconceptions by using LSA in the match processes. The tutor's dialog moves are formulated by the DAN in conjunction with LSA and the curriculum script.

When each student's turn is processed, mechanisms are needed to parse and segment the natural language input, to assign speech acts and clauses to discourse categories, to evaluate the quality of the student's contributions, and to formulate the dialog moves of the tutor. Researchers have proposed a number of category systems for classifying and marking up the speech acts in the tutorial dialog (Core, Moore, & Zinn, 2000; Person et al., 2001; Shah, Evens, Michael, & Rovick, 2002). Some of these analytical systems require expert human judges to segment, mark up, and classify the constituents, whereas other systems have most or all of these components automated. In the case of AutoTutor, all of the components are automated. Manual coding of over 100 hours of human to human tutoring sessions was done to determine the tutoring strategies and discourse patterns of untrained human tutors (Graesser & Person, 1994; Person, Graesser, Magliano, & Kreuz, 1994). The five step dialogue frame and the dialogue moves identified in this analysis were used in the development of AutoTutor. The log files reflect the components of the tutorial dialogue that were identified as important for supporting the dialogue management and in evaluating of system's effectiveness. In the future the log files will be analyzed to determine the relationship between discourse patterns and learning gains.

Transcripts and log files of the tutoring sessions record any aspect of the tutoring session that will be of further help in evaluating the quality of the tutor-student interactions, and the computational architecture of the system. The log files in AutoTutor record the student-tutor dialogue, syntactic parses of the student's turns, the categories of the speech act classifier, states of the DAN, and LSA evaluations of the student's input. The remainder of this paper discussed the contents of the log files as they are currently implemented in AutoTutor.

Contents of the Log Files

We are currently conducting a large-scale experiment on the tutoring of conceptual physics at three universities (University of Pittsburgh, Rhodes College and the University of Memphis). All of the tutoring systems present conceptual physics problems such as the one below.

> 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? Explain.

It should be noted that a good answer would require about a paragraph of information because the student is expected to explain his/her reasoning. In this section we will examine a section of one of the intelligent tutoring systems, called Why/AutoTutor. Why/AutoTutor's log file was collected during one of these physics tutoring sessions. As we discuss different aspects of the log files, we will show an excerpt from the log file. The Appendix contains the complete section of the subtopic being discussed.

Each session begins by recording the subtopic being tutored, the level of difficulty (easy, moderate, hard), and the time and date the session began. In the section of the log file we discuss here, the topic is "neck-injury" and the session has a low level of difficulty.

```
TUTOR-TOP-4>  Neck-injury
TUTOR-START-0> SubTopic started: Monday, March 18, 2002 2:56:13 PM GMT+06:00
TUTOR-SUB-4> HW-4-QA-Easy
```

An important component of a dialogue management system is the classification of the input to determine the user's intent (Allen, Byron, Dzikovska, Ferguson, Galescu & Stent, 2001). Dialogue acts can have multiple levels of meaning. One can ask "Can you repeat the question?", assert "I would like you to repeat the question.", or issue a command "Repeat the question." The speaker's intent is the same, but they are different speech acts with different surface forms. Various mark-up languages have been developed to classify the user's input into categories that are useful to dialogue management systems in particular domains. DAMSL (Dialogue Act Mark up in Several Layers) is a classification scheme developed by the analysis of advice-giving and task-oriented dialogues (Allen & Core, 1997). This classification scheme includes categories for accepting and rejecting proposals that are specific to the types of dialogues analyzed. Shah, et al (2002) have proposed another classification scheme based upon an analysis of computer-based human to human tutorial dialogue. Student initiatives, defined as any student contribution that is not an answer to a tutor's question, are categorized along four dimensions: communicative goal, surface form, focus, and hedging. In tutorial dialogue it is necessary to appropriately classify the student's input. If the system is not responsive to the student's directives and questions, then they quickly become frustrated and limit how they interact with the system.

In order to capture the user's intent (at least to some extent) and to encourage mixed-initiative dialog, AutoTutor uses a speech act classifier that categorizes speech acts in a student's turn into one of 24 possible categories, including question categories, and metacommunicative and metacognitive frozen expressions (see Louwerse, Olney, Mathews, Marineau & Mitchell, submitted; Olney, Louwerse, Mathews, Marineau, Mitchell, submitted). The questions are classified into 16 categories based on a taxonomy developed by Graesser, Person, and Huber (1992). The classifier uses both a syntactic and a semantic analysis for this purpose, and correctly classifies questions with 86% accuracy (Olney et al, submitted). Accuracy ratings for each question category are provided in Table 1. The speech act classifier for AutoTutor was developed to provide speech act classification at a more general level. As we develop AutoTutor to become more responsive to student's questions, we needed a mechanism that would accurately identify the type of question the student was asking without being restricted to a particular domain. The log files contain the parser output as well as the speech act category ("contribution" in the example below).

Table 1. Question Classifier Accuracy.

Question Category	*Accuracy Rating*
Verification	96%
Comparison	95%
Disjunctive	92%
Concept Completion	84%
Definition	66%
Example	95%
Interpretation	54%
Feature Specification	82%
Quantification	94%
Causal Antecedent	66%
Causal Consequence	81%
Goal Orientation	92%
Enablement	87%
Instrumental	76%
Expectational	83%
Judgmental	84%

In the log file below AutoTutor first uses one of a large variety of canned expressions classified according to their use in dialog (prompt-marker, elaboration-marker, hint-marker, etc.). It then introduces a context and asks a question. The time it takes AutoTutor to respond and the time of the student to respond is presented next, followed by the student input. The category in which the student's answer is classified is given next as well as a parse of the student answer, used in determining the speech act category.

```
TUTOR-DM-147> Let's talk about something else.
TUTOR-DM-148>  \atcmessage:question\ 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?
Explain.\/atcmessage:question\
AT-TRT-UTTERANCE-TIME-54> 21.6320000000019
AT-SRT-UTTERANCE-TIME-54> 144.946999999997
STUDENT-55> The passengers get neck injuries because the car behind them is
applying a force in one direction and there is nothing to absorb the force of
the head snapping back
TUTOR-SAC-53> CONTRIBUTION
PARSE-53> (S (NPL (DT The) (NNS passengers)) (VP (VBP get) (NP (NPL (NN neck)
(NNS injuries)) (SBAR (WHNP (IN because)) (S (NP (NPL (DT the) (NN car)) (PP
(IN behind) (NPL (PRP them)))) (VP (VP (VBZ is) (VP (VBG applying) (NPL (DT a)
(NN force)) (PP (IN in) (NPL (CD one) (NN direction))))) (CC and) (VP (ADVP
(RB there)) (VBZ is) (NP (NN nothing) (TOINF (VP (TO to) (VP (VB absorb) (NP
(NPL (DT the) (NN force)) (PP (IN of) (NPL (DT the) (NN head) (VBG snapping)
(NN back))))))))))))))
```

In addition to each student and tutor turn, the log files provide a continuous dialogue history of the student's contributions used in calculating the subtopic coverage score. The current contribution is the student's current input to be analyzed. Since this is the first contribution of the student, the current contribution is the same as the dialog history. The dialog history can get quite long of course.

```
DIALOG-HISTORY-48>  the passengers get neck injuries because the car behind
them is applying a force in one direction and there is nothing to absorb the
force of the head snapping back
CURRENT-CONTRIBUTION-48> the passengers get neck injuries because the car
behind them is applying a force in one direction and there is nothing to
absorb the force of the head snapping back
```

Using the dialogue history and the student's current contribution, the log file records the extent to which this information covers expectations and misconceptions that are anticipated in the curriculum script. These coverage scores may correspond to global and local spans of information. LSA cosine metrics (which vary from 0 to 1) assess the extent to which the entire dialog history and the immediate local contribution cover particular bad answers (called “badPoints”), and expected good answers (called “goodPoints”). There are other coverage metrics, which need not be addressed, in the present context. There is a completeness measure that assesses the proportion of expectations that are covered (i.e., the cosine metric exceeds some threshold). The verbosity measure is an index of how wordy the student has been throughout the previous session.

```
DIALOG-HISTORY-ASSESSMENTS-48>
TUTOR-ASR-95> misconceptionAnswer: 0.59989697
TUTOR-ASR-95> subTopicFirstMove: 1.0
TUTOR-ASR-95> topicCoverage: 0.38809592
TUTOR-ASR-95> scriptCoverage: 1.0
TUTOR-ASR-95> completeness: 0.25
TUTOR-ASR-95> badPoints(4): 0.36137545
TUTOR-ASR-95> badPoints(3): 0.23158917
TUTOR-ASR-95> badPoints(2): 0.7409022
TUTOR-ASR-95> badPoints(1): 0.6127475
TUTOR-ASR-95> subTopicCoverage: 0.25
TUTOR-ASR-95> badAnswer: 0.7409022
```

```
TUTOR-ASR-95> verbosity: 1.0
TUTOR-ASR-95> goodPoints(4): 0.55672544
TUTOR-ASR-95> goodPoints(3): 0.6629203
TUTOR-ASR-95> goodPoints(2): 0.22583447
TUTOR-ASR-95> goodPoints(1): 0.49844313
TUTOR-ASR-95> topicAbility: 0.4932672
TUTOR-ASR-95> scriptAbility: 0.6085928
TUTOR-ASR-95> goodAnswer: 0.7571305
TUTOR-ASR-95> misconceptionPoints(1): 0.59989697
TUTOR-ASR-95> subTopicAbility: 0.4932672
TUTOR-ASR-95> ------------------------
LOCAL-ASSESSMENTS-48>
TUTOR-ASR-96> misconceptionAnswer: 0.59989697
TUTOR-ASR-96> subTopicFirstMove: 0.0
TUTOR-ASR-96> topicCoverage: 0.2966336
TUTOR-ASR-96> scriptCoverage: 1.0
TUTOR-ASR-96> completeness: 0.25
TUTOR-ASR-96> badPoints(4): 0.36137545
TUTOR-ASR-96> badPoints(3): 0.23158917
TUTOR-ASR-96> badPoints(2): 0.7409022
TUTOR-ASR-96> badPoints(1): 0.6127475
TUTOR-ASR-96> subTopicCoverage: 0.25
TUTOR-ASR-96> badAnswer: 0.7409022
TUTOR-ASR-96> verbosity: 1.0
TUTOR-ASR-96> goodPoints(4): 0.55672544
TUTOR-ASR-96> goodPoints(3): 0.6629203
TUTOR-ASR-96> goodPoints(2): 0.22583447
TUTOR-ASR-96> goodPoints(1): 0.49844313
TUTOR-ASR-96> topicAbility: 0.49326718
TUTOR-ASR-96> scriptAbility: 0.60769707
TUTOR-ASR-96> goodAnswer: 0.7571305
TUTOR-ASR-96> misconceptionPoints(1): 0.59989697
TUTOR-ASR-96> subTopicAbility: 0.49326718
TUTOR-ASR-96> ------------------------
```

Based on the good and bad answer scores calculated for the student's current contribution, the program provides mean and z-scores for the good and bad aspects. Using these scores the student is provided with positive-positive, positive-neutral, neutral-neutral, neutral-negative or negative-negative feedback. The tutor uses these categories of feedback for generating the appropriate discourse markers (e.g. excellent, good, okay, I see, not really, nonsense).

```
FEEDBACK-1> Max GoodAspect Z-score -1.6601369582882215
FEEDBACK-1> Max BadAspect Z-score -0.2899272812341066
FEEDBACK-1> GoodAspectMean 0.1514837167416086
FEEDBACK-1> BadAspectMean 0.14083064643548063
TUTOR-DMT-190> neutralNeutralFeedback Aspect = 0 Number = 0
TUTOR-DM-157> okay \pau=700\
```

Each expectation (aspect) of the ideal answer has hints, prompts, and elaborations (assertions) that are written in the curriculum scripts (Graesser, et al., 1999). After selecting the appropriate type of dialogue move for AutoTutor to use next, the log files display a selection of the particular dialogue move chosen, and the LSA cosine score that indicates which of the moves would best increase the aspect coverage if the student answers correctly. The system then records the aspect and dialogue move the tutor uses next.

```
TUTOR-DMT-182> prompt Aspect = 0 Number = 0
CHOOSE-PROMPT-13> (1) value: 0.43479836 they need the same \emp\ acceleration.
CHOOSE-PROMPT-13> (3) value: 0.4981563 as the passenger's \emp\ body.
CHOOSE-PROMPT-13> (2) value: 0.18854009 moving at the same \emp\ rate.
TUTOR-DM-151>  What about this? Since the they are connected by the neck, the
passenger's head needs to have the same acceleration as the passenger's
__________?
STUDENT-56> body
```

Once a student has provided the correct answer to one expectation (aspect) of the ideal answer, the log files show the expectation rankings by which the system chooses the next expectation to address. LSA coverage scores are shown for the expectations that have not yet been fully answered (sub-threshold scores). The coherence score is an LSA score ranking each expectation by how well it would contribute to the overall coherence of the dialogue, i.e., overlap with the previous expectation that was just covered. The redundancy score indicates how much the coverage of one particular expectation would cover other expectations in the ideal answer.

```
Subthreshold Ranks___
   Aspect (4)  In an attempt to produce the required large force, the neck
gets stretched and may get injured damaging its muscles and ligaments.
0.55672544
   Aspect (1)  When a car is struck from behind the force of impact will cause
a large forward acceleration of the car. 0.49844313

Coherence  Ranks___
   Aspect (4)  In an attempt to produce the required large force, the neck
gets stretched and may get injured damaging its muscles and ligaments.
0.55672544
   Aspect (1)  When a car is struck from behind the force of impact will cause
a large forward acceleration of the car. 0.49844313

Redundancy  Ranks___
   Aspect (4)  In an attempt to produce the required large force, the neck
gets stretched and may get injured damaging its muscles and ligaments.
0.55672544
   Aspect (1)  When a car is struck from behind the force of impact will cause
a large forward acceleration of the car. 0.49844313
```

ASPECT-CHANGE-11> aspect 4 started: Monday, March 18, 2002 2:59:36 PM

The student provided two correct expectations in the ideal answer. Aspect (4) was chosen as the answer aspect that would assist the student in the completion of the ideal answer. After each aspect change, the tutor re-evaluates and generates the dialogue move most likely to lead the student toward the ideal answer.

Discussion

Log files are used in AutoTutor for keeping records of tutoring sessions, with the hope of incrementally improving the system's performance. Analysis of AutoTutor's log files from sessions in computer literacy provided useful information that was used to improve the quality of the dialogue and the functioning of LSA in the most recent version of AutoTutor. In the initial analysis it was noted that the limited number of meta-cognitive and meta-communicative statements that the tutor offered made his conversational style redundant and less natural. In later versions of AutoTutor, we have expanded the number and variety of meta-cognitive and meta-communicative statements offered by the tutor. Another deficit discovered in this analysis was that the users were unsure of when it was their turn to respond to the tutor. If the tutor's final dialogue move was an elaboration (assertion), the student was unclear how he/she was expected to respond (Person, Graesser, Harter, Mathews, and the TRG, 2000). The tutor's final dialogue move was changed so he would offer a hint, a prompt, or a summary. This change alleviated some of the turn-taking confusion.

Analysis of the LSA scores documented in the log files indicated that there were adjustments that could be made to the curriculum scripts and the LSA space to improve AutoTutor's performance. The LSA values of the students' input were analyzed to assess LSA's ability to accurately evaluate the student's knowledge. A student's answer that used content words contained within the curriculum scripts would artificially inflate the LSA score. The use of important concept words would result in LSA giving the answer a high rating even though it contained a misconception. The curriculum scripts were rewritten to contain generic forms of important content words. These generic concepts were then associated with the more specific terms in the concept list. This prevented high LSA scores for answers that contained the right words but the wrong context. The evaluation of LSA led to the question of what information should be included when training LSA. Would a larger training corpus lead to better performance? What would happen to performance if the corpus contained historical information, or examples of the misconceptions a student might have? Five different physics corpora were analyzed to determine what type of corpora would provide the best LSA performance (Olde, Franceschetti, Karnavat, Graesser, & the TRG, in press). The corpora differed in size and the amount of historical information and misconceptions they contained. The study showed that the size of the corpus had little impact on LSA performance, and a small corpus of relevant material could provide acceptable performance. There was also no benefit in removing irrelevant information, such as historical facts and examples of misconceptions and errors.

A number of markup initiatives have been introduced in the past (Allen & Core, 1997; Carletta, Isard, Isard, Kowtko, Newlands, Doherty-Sneddon, & Anderson, 1997; Di Eugenio, Jordan & Pylkkänen, 1997). The purpose of these markup initiatives is to provide a corpus with a detailed tag set that would facilitate subsequent discourse analyses. These initiatives are extremely useful in many areas of language and dialog analysis. Furthermore, they provide a level of detail that can currently not be achieved by any automatic tagging system. On the down side, these initiatives are extremely labor intensive and time consuming. Automatic tagging is less detailed, but adequate for the purposes of many research projects. It remains to be seen whether Why/AutoTutor’s current log file will go the distance in tracking and improving the tutorial dialog, but the log files from the earlier versions of AutoTutor have provided us with valuable information used to improve the conversational quality and pedagogical effectiveness of the current AutoTutor.

Acknowledgements

This research was supported by grants from DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (N00014-00-1-0600) and the National Science Foundation (SBR 9720314 and REC 0106965). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ONR or NSF.

References

Aleven, V., Popescu, O., & Koedinger, K.R. (2001). Towards tutorial dialog to support self-explanation: Adding Natural language understanding to a cognitive tutor. In J.D. Moore, C. Luckhardt, & W.L. Johnson (Eds.), Artificial intelligence in education (pp. 246-255). Amsterdam: IOS Press.

Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L. & Stent, A. (2000). Towards a generic dialogue shell, Natural Language Engineering,6(3), 1-16.

Allen, J. & Core, M. (1997). Coding dialogs with the DAMSL annotation scheme. In working notes of the AAAI Fall 1997 Symposium on Communicative Action in Humans and Machines.

Carletta, J., Isard, A.., Isard, S., Kowtko, J., Newlands, A., Doherty-Sneddon, G., Anderson, A. (1997). The reliability of a dialogue structure coding scheme. Computational Linguistics, 23, 13-3.

.
Corbett, A., Anderson, J., Graesser, A., Koedinger, K. & VanLehn, K. (1999). Third generation computer tutors: learn from or ignore human tutors? In Proceedings of the 1999 Conference of Computer-Human Interaction, (pp 85-86). New York: ACM Press.

Core, M., Moore, J., & Zinn, C.(2000). Supporting constructive learning with a feedback planner. In Building Dialogue Systems for Tutorial Applications. Papers from the 2000 AAAI Fall Symposium. (pp. 1-9). Cape Cod, MA: AAAI Press.

Di Eugenio, B., Jordan, P.W., & Pylkkänen, P. (1997). The COCONUT project: dialogue annotation manual. Unpublished manuscript.

Franceschetti, D.R., Karnavat, A., Marineau, J., McCallie, G.L., Olde, B.A., Terry, B.L., & Graesser, A.C. (2001). Development of physics text corpora for latent semantic analysis. In Proceedings of the 23rd Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Freeman, R.K. (2000). Plan-based dialog management in a physics tutor. Proceedings of the 6th Applied Natural Language Conference.

Graesser, A.C., Person, N., Harter, D., & TRG (2001). Teaching tactics and dialog in AutoTutor. International Journal of Artificial Intelligence in Education, 12, xxx-xxx.

Graesser, A. C., Person N., & Huber, J. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, and A. Graesser, (Eds.), *Questions and information systems*. Earlbaum, Hillsdale, NJ.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied Cognitive Psychology, 9, 359.1-28.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. AI Magazine, 22, 39-51.

Graesser, A.C., Wiemer-Hastings, P., Weimer-Hastings, K., Kreuz, R.J., & the TRG (1999). AutoTutor: A simulation of a human tutor. Cognitive Systems Research, 1, 35-51.

Graesser, A.C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N.K., & the TRG (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. Interactive Learning Environments, 8, 129-148.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquistion, induction, and representation of knowledge. Psychological Review,104, 211-240.

Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

Louwerse, M., Olney, A., Mathews, E., Marineau, J., Mitchell, H.H. (2002). Toward an automated speech act classifier: Implications for intelligent tutoring systems. Unpublished manuscript.

Moore, J.D. (1995). Participating in explanatory dialogues. Cambridge, MA: MIT Press.

Olde, B. A., Franceschetti, D., Karnavat, A., & Graesser, A. C. (in press). The right stuff: Do you need to sanitize your corpus when using latent semantic analysis. Proceedings of the 23rd Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Olney, A., Louwerse, M., Mathews, E., Marineau, J., Mitchell, H.H. (2000). Surface cues for categorizing conceptual questions. Submitted to

Person, N.K., Graesser, A.C., Harter, D., Mathews, E., & TRG (2000). Dialogue move generation and conversation management in AutoTutor. In Building Dialogue Systems for Tutorial Applications, Papers from the 2000 AAAI Fall Symposium. (pp. 45-51). Cape Cod, MA: AAAI Press.

Person, N.K., Graesser, A.C., Kreuz, R.J., Pomeroy, V., & TRG (2001). Simulating human tutor dialog moves in AutoTutor. International Journal of Artificial Intelligence in Education, 12, 23-39.

Person, N.K., Graesser, A.C., Magliano, J. P., & Kreuz, R.J. (1994). Inferring what the student knows in one-to-one tutoring: The role of student questions and answers. Learning and Individual Differences, 6, (2), 205-229.

Rose, C.P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., & Weinstein, A. (2001). Interactive conceptual tutoring in Atlas-Andes. In J.D. Moore, C. Luckhardt, & W.L. Johnson (Eds.), Artificial Iintelligence in education (pp. 257-266). Amsterdam: IOS Press.

Shah, F., Evens, M., Michael, J., & Rovick, A. (2002). Classifying student initiatives and tutor responses in human keyboard-to-keyboard tutoring sessions. Discourse Processes, 33, (1), 23-52.