

## Simulating Smooth Tutorial Dialogue with Pedagogical Value

Art Graesser, Stan Franklin, Peter Wiemer-Hastings,  
and the Tutoring Research Group

The University of Memphis  
Department of Psychology  
Campus Box 526400  
Memphis TN 38152-6400  
a-graesser@memphis.edu

### Abstract

This paper describes the design of AutoTutor, a fully automated computer tutor that simulates dialogue moves of normal human tutors and that will eventually incorporate sophisticated tutoring strategies. AutoTutor follows a curriculum script that presents subtopics in different forms, such as didactic content, graphic displays, questions, and problems. The tutor selects dialogue moves that assist the learner in the active construction of knowledge as the learner answers questions or solves problems posed by the tutor. The goal is to have the tutor produce dialogue moves that fit the conversational context, that are sensitive to the learner's abilities, and that have pedagogical value. The dialogue moves include immediate feedback, pumping, prompting, splicing, hinting, requestioning, and summarizing. The tutor's moves are delivered by a talking head with appropriate facial expressions and synthesized speech. The first version of AutoTutor should be completed in the spring of 1998 on the topic of computer literacy.

### Can a Computer be a Good Partner in Tutorial Dialogue?<sup>1</sup>

Designers of intelligent tutoring systems (ITS) have frequently had the vision of a fully automated computer tutor that trains students on skills and domain knowledge. Unfortunately, language and discourse have constituted serious barriers. As a consequence, language and discourse facilities have been either nonexistent or extremely limited in even the most impressive and successful intelligent tutoring systems available, such as Anderson's tutors for geometry, algebra, and computer languages (Anderson, Corbett, Koedinger, & Pelletier, 1995), Van Lehn's tutor for basic mathematics (Van Lehn, 1990), and Lesgold's tutor for diagnosing and repairing electronic equipment (Lesgold, Lajoie, Bunzo, & Eggan, 1992). There have been some attempts to augment ITS's with language and dialogue facilities (Holland, Kaplan, & Sams, 1995; Moore, 1994). But such attempts have been limited by (1) the inherent difficulty of getting the computer to "compre-

hend" the language of users, including utterances that are not well-formed syntactically and semantically, (2) the difficulty of getting computers to effectively use a large body of open-ended, fragmentary world knowledge, and (3) the lack of research on human tutorial dialog and on patterns of normal discourse. These difficulties have been exacerbated by insufficient communication between the fields of discourse processing, education, computational linguistics, and ITS developers.

Advances in research during the last five years make it much more feasible to develop a computer tutor that simulates "smooth" tutorial dialog, i.e., speech acts that are appropriate in the context of the conversation. These recent developments have provided approximate solutions to the above three major barriers. The tutoring system that we have been developing, called AutoTutor, incorporates these approximate solutions in addition to more established computational procedures.

AutoTutor attempts to "comprehend" text that the learner types into the keyboard and to formulate appropriate discourse contributions. The generated discourse contributions may be in the form of printed text, synthesized speech, graphic displays, animation, and simulated facial expressions. That is, the tutor will speak in different media. However, the primary technological contribution of our tutor lies in formulating helpful discourse contributions, as opposed to generating a fancy display of interface features. Simply put, our goal is to determine "what the tutor should say next" (i.e., the conceptual content), not "how the tutor should say it" (i.e., in digitized speech, synthesized speech, print, versus a talking head). There eventually will be different versions of AutoTutor in an effort to simulate (a) skilled and unskilled human tutors who vary in domain expertise and tutoring experience and (b) ideal tutoring strategies that have been identified in the field of education and by developers of intelligent tutoring systems. However, the first version of AutoTutor (completed in the spring of 1998) simulates the dialogue moves of unskilled human tutors.

The feasibility of a computer tutor is fortified by our previous research on human tutoring that analyzed videotapes of approximately 100 hours of untrained tutors in naturalistic tutoring sessions (Graesser, Person,

Copyright ©1998, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

& Magliano, 1995; Person, Kreuz, Zwaan, & Graesser, 1995). Approximately 95% percent of the tutors in actual school systems are untrained. They have moderate domain knowledge and no training in tutoring skills. These tutors are, however, extremely effective; they enhance learning by 0.4 to 2.3 standard deviation units compared to classroom controls (Cohen, Kulik, & Kulik, 1982; Bloom, 1984). This result is compatible with the claim that there is something about interactive discourse that is responsible for learning gains. Our previous research revealed that human tutors and learners have a remarkably incomplete understanding of each other's knowledge base and that many of each other's contributions are not deeply understood. Breakdowns in communication occur to the extent that the "common ground" (i.e., shared knowledge) of the speech participants decreases. It is not fine-tuned "student modeling" that is important, but rather a tutor that serves as a conversation partner when common ground is minimal. We also discovered that a key feature of effective tutoring lies in generating discourse contributions that assist learners in the active construction of subjective explanations, elaborations, and mental models of the material. Other researchers have also proposed that active, subjective constructions of explanations are critical for learning, and have a greater impact than merely presenting information to learners (Bransford, Goldman, & Vye, 1991; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & La Vaucher, 1994). Human tutors assist this subjective construction of knowledge by delivering collaborative discourse moves. Our analyses of naturalistic tutoring revealed that the tutor's discourse moves include questioning, requestioning, immediate feedback, pumping, prompting, hinting, summarizing, splicing in correct information, and revoicing student contributions. AutoTutor is designed to simulate these tutoring moves in a manner that is pragmatically smooth and pedagogically effective.

Researchers in education and ITS have identified a number of ideal tutoring strategies, such as: the Socratic method (Collins, 1985), modeling-scaffolding-fading (Collins, Brown, & Newman, 1989), reciprocal training (Palincsar & Brown, 1984), anchored learning (Bransford, Goldman, & Vye, 1991), error diagnosis and correction (Anderson et al., 1995; van Lehn, 1990; Lesgold et al., 1992), frontier learning (Sleeman & Brown, 1982), building on prerequisites (Gagné, 1977), and sophisticated motivational techniques (Lepper, Aspinwall, Munne, & Chabay, 1990). Detailed discourse analyses have been performed on samples of these sophisticated tutoring strategies (Fox, 1993; Hume, Michael, Rovick, & Evens, 1996; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Ranney, & Trafton, 1992; Putnam, 1987). It should be noted, however, that these sophisticated tutoring strategies were practically nonexistent in the unskilled tutoring sessions that we videotaped and analyzed (Graesser et al., 1995). Later versions of AutoTutor will incorporate some of

these ideal tutoring strategies, but the first version simulates the dialogue moves of the normal unskilled tutor, which are known to be very effective according to previous research (Cohen et al., 1982).

## Components of AutoTutor

AutoTutor incorporates both classical symbolic architectures (e.g., those with propositional representations, conceptual structures, and production rules) and architectures with multiple soft constraints (i.e., neural network models, fuzzy descriptions and controllers, and latent semantic analysis). The major components of this system are briefly specified below.

### Curriculum Script

The macrostructure that guides the tutorial dialog consists of a curriculum script (McArthur et al., 1991; Putnam, 1987) with didactic descriptions, tutor-posed questions, cases, problems, figures, and diagrams (along with good responses to each subtopic). AutoTutor currently has a curriculum script for the domain of computer literacy. The script includes three topics (hardware, operating system, and internet) and 12 subtopics per topic. The subtopics vary in difficulty and have one of four formats: Question + Answer, Problem + Solution, Didactic information + Question + Answer, and Graphic display + Question + Answer. The Answer or Solution content that is associated with each subtopic include the following data structures:

- an ideal answer
- a list of different good answers
- a list of different bad answers
- a list of hints, scaled on directness and difficulty
- a succinct summary of the answer
- a list of anticipated student questions and answers to these questions
- a verbal description of the graphic display (if there is a display)

Subtopics are selected by production rules that are sensitive to the learner's abilities, initiative, and other global parameters. The learner's ability level is computed by the quality of student contributions (i.e., answers, solutions) during the exchanges in previous subtopics in the session.

### Language Modules

Language modules analyze the content of the message that the learner types in from the keyboard during a particular conversational turn. The sequence of words and punctuation marks in a turn are first segmented into speech act units, which are then classified into speech act categories (i.e., Question, Contribution, versus Short Response). The language modules include a lexicon (Francis & Kučera, 1982; Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), a connectionist network that identifies the syntactic classes of words, a

dictionary of frozen expressions, software agents implemented as codelets that sense surface linguistic features (Franklin, Graesser, Olde, Song, & Negatu, 1996; Zhang, Franklin, Olde, Wan, & Graesser, 1998), and a recurrent connectionist network that formulates predictions about the next speech act category (Graesser, Swamer, & Hu, 1997). For example, in a simple version of AutoTutor, a Question is signaled by a question mark, a Short Response is identified if there is match with a list of short responses (e.g., "yes," "okay"), and all other speech acts are classified as Contributions.

### Latent Semantic Analysis

Those speech acts of the learner that are classified as Contributions (as opposed to Questions and Short Responses) are analyzed on various dimensions of quality. Latent semantic analysis (LSA, Landauer & Dumais, 1997) provides the backbone for representing the world knowledge that is needed to make these evaluations. LSA reduces a large corpus of texts to a 100–300 dimensional space through a statistical method called singular value decomposition. LSA has had remarkable success in grading essays of students and answering questions on multiple choice tests. In fact, the performance of LSA has been nearly equivalent to human performance on a number of tasks and measures (Landauer & Dumais, 1997). The LSA space for the domain of computer literacy is based on two books on computer literacy and approximately 20 articles that focus on hardware, operating systems, and the internet. The LSA space is used to compute the relatedness (i.e., similarity in content) of any two sets of words. Such relatedness scores are extremely useful because they permit us to compute the truth, relevance, and other dimensions of the quality of student contributions. For example, the relevance of student contribution C to the subtopic S is the maximum relatedness score between the set of words in C and the set of words in each description (i.e., sentence, content specification) within S. Truth is computed as the maximum relatedness score between C and all descriptions in the entire set of 36 subtopics. There are other dimensions of answer quality in addition to truth and relevance. A very important feature of LSA is that it can be used on speech acts and texts that are *not* syntactically and semantically well-formed.

### Tutor Dialogue Moves

After the student types in the content of his or her turn, the tutor generates a dialogue move in one of several categories, which are illustrated below.

- Positive immediate feedback: "That's right," "Yeah"
- Neutral immediate feedback: "Okay," "Uh-huh"
- Negative immediate feedback: "Not quite," "No"
- Pumping for more information: "Uh-huh," "What else"
- Prompting for specific information: "The memories of the CPU are ROM and \_\_\_\_\_"

- Hinting: "The hard disk can be used for storage"
- Splicing in correct content after a student error
- Requestioning: "So once again, what is the function of a CPU?"
- Summarizing: "So to recap," <succinct recap of answer to question>

It should be noted that a tutor turn may contain two dialogue moves, such as neutral immediate feedback followed by a hint.

The *categories* of dialogue moves during a turn are determined by a set of fuzzy production rules. These rules are tuned to:

- the truth, relevance, and other measures of the quality of the student's recent contribution
- global parameters that refer to the ability and initiative of the student
- phases during the subtopic, topic, and tutoring session
- the style and skill of the tutor

The content of the selected categories are formulated by production rules that select descriptions from the curriculum script (e.g., the hint list, the good answer list) and from lists of frozen expressions (e.g., lists of words denoting neutral feedback).

### Talking head

Most of the tutor's dialogue moves are delivered by a talking head that is synchronized with synthesized speech (Cohen & Massaro, 1994). The facial expressions and intonation in the immediate feedback are sensitive to the truth, relevance, and quality of the student's most recent contribution. The parameters of the facial expressions and intonation are generated by fuzzy production rules.

### Student questions

AutoTutor also can handle student questions that may occur during the collaborative dialogue between tutor and learner for a subtopic. However, studies of naturalistic tutoring have revealed that student questions are not very frequent during tutoring and that the questions that are asked cover a limited set of question categories (Graesser & Person, 1994). AutoTutor currently anticipates and can answer definitional questions ("What does X mean?"), which is the most frequent category of question that students ask (Graesser & Person, 1994). Therefore, AutoTutor to some extent provides a mixed-initiative dialog between the tutor and student.

### Evaluating AutoTutor

The quality of the tutorial dialog generated by AutoTutor is currently being evaluated. Its success will be tested in several ways. The fidelity of particular language modules will be evaluated with respect to recall,

precision, and other performance measures used by researchers in computational linguistics (DARPA, 1995; Lehnert, 1997). Experts in discourse and education will evaluate the appropriateness, relevance, and pedagogical quality of the dialog contributions generated by the computer tutor. Turing tests will be performed at a fine-grained level in order to assess whether the learner (or a neutral human observer) can discriminate whether particular dialog moves are generated by the computer versus a human tutor. There eventually will be assessments of whether AutoTutor produces significant learning gains, compared to control conditions. However, our immediate concerns are with the fine-grained discourse contributions of AutoTutor.

### Acknowledgments

This research was funded by the National Science Foundation (SBR 9720314). The members of the Tutoring Research Group are Ashraf Anwar, Myles Bogner, Tim Brogdon, Patrick Chipman, Scotty Craig, Stan Franklin, Max Garzon, Barry Gholson, Art Graesser, Doug Hacker, Derek Harter, Jim Hoeffner, Xiangen Hu, Bianca Klettke, Roger Kreuz, Kristen Link, William Marks, Lee McCaulley, Fergus Nolan, Brent Olde, Natalie Person, Victoria Pomeroy, Melissa Ring, Charles Webster, Matt Weeks, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Holly Yetman, and Zhaohua Zhang.

### References

- Anderson, J., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4, 167-207.
- Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.
- Bransford, J., Goldman, S., & Vye, N. (1991). Making a difference in people's ability to think: Reflections on a decade of work and some hopes for the future. In R. Sternberg & L. Okagaki (Eds.), *Influences on children* (pp. 147-180). Hillsdale, NJ: Erlbaum.
- Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Chi, M., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.
- Cohen, M., & Massaro, D. (1994). Development and experimentation with synthetic visible speech. *Behavior Research Methods, Instruments, and Computers*, 26, 260-265.
- Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19, 237-248.
- Collins, A. (1985). Teaching reasoning skills. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol.2, pp. 579-586). Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J., & Newman, S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- DARPA (1995). Proceedings of the Sixth Message Understanding Conference (MUC-6). San Francisco: Morgan Kaufman Publishers.
- Fox, B. (1993). *The Human Tutoring Dialogue Project*. Hillsdale, NJ: Erlbaum.
- Francis, W., & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin Company.
- Franklin, S., Graesser, A., Olde, B., Song, H., & Negatu, A. (1996). Virtual Mattie: An intelligent clerical agent. AAAI Symposium on Embodied Cognition and Action, Cambridge, MA.
- Gagné, R. (1977). *The conditions of learning* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Goldman, S., Pellegrino, J. & Bransford, J. (1993). Assessing programs that invite thinking. In H. O'Neill & E. Baker (Eds.), *Technology assessment: Estimating the future*. Hillsdale, NJ: Erlbaum.
- Graesser, A., & Person, N. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.
- Graesser, A., Person, N., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology*, 9, 359-387.
- Graesser, A., Swamer, S., & Hu, X. (1997). Quantitative discourse psychology. *Discourse Processes*, 23, 229-263.
- Hobbs, J., Appelt, D., Tyson, M., Bear, J., & Israel, D. (1992). SRI International: Description of the FASTUS System used for MUC-4. Proceedings of the Fourth Message Understanding Conference (MUC-4) (pp. 268-275). San Mateo, CA: Morgan Kaufmann.
- Holland, V., Kaplan, J., & Sams, M. (1995) (Eds.). *Intelligent language tutors*. Mahwah, NJ: Erlbaum.
- Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *The Journal of the Learning Sciences*, 5, 23-47.
- Jacobs, P. (1992) (Ed.). *Text-based intelligent systems: Current research and practice in information extraction and retrieval*. Hillsdale, NJ: Erlbaum.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lehnert, W. (1997). Information extraction: What have we learned? *Discourse Processes*, 23, 441-470.

- Lepper, M., Aspinwall, L., Mumme, D., & Chabay, R. (1990). Self-perception and social-perception processes in tutoring: Subtle social control strategies of expert tutors. In J. Olson & M. Zanna (Eds.), *Self-inference processes: The Ontario symposium* (pp. 217-237). Hillsdale, NJ: Erlbaum.
- Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin & R. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems* (pp. 201-238). Hillsdale, NJ: Erlbaum.
- Massaro, D., & Cohen, M. (1995). Perceiving talking faces. *Psychological Science*, 4, 104-109.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197-244.
- Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*, 2, 277-305.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Five papers on WordNet. Cognitive Science Laboratory, Princeton University, No. 43.
- Moore, J. (1994). *Participating in explanatory dialogues: Interpreting and responding to questions in context*. Cambridge, MA: The MIT Press.
- Palinscar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, 1, 117-175.
- Person, N., Kreuz, R., Zwaan, R., & Graesser, A. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13, 161-188.
- Putnam, R. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal*, 24, 13-48.
- VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.
- Zhang, Z., Franklin, S., Olde, B., Wan, Y., & Graesser, A. (1998). Natural language sensing for autonomous agents. IEEE Proceedings on Intelligence and Systems.