

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Arthur C. Graesser , Fazel Keshtkar , and Haiying Li

The Oxford Handbook of Language and Social Psychology

Edited by Thomas M. Holtgraves

Print Publication Date: Sep 2014 Subject: Psychology, Social Psychology

Online Publication Date: Jan 2014 DOI: 10.1093/oxfordhb/9780199838639.013.003

Abstract and Keywords

Computer learning environments of the future need to be sensitive to the social, cognitive, emotional, and motivational (SCEM) states of students as they learn in their social environment. Language and discourse plays a central role in tracking SCEM states and influencing how the computer responds to promote learning. This essay describes a number of computer tutors that are sensitive to these psychological factors and thereby help students learn. Computer agents are central to the design of these systems. These systems include one-on-one tutoring, conversational dialogues (a tutor agent and student agent conversing with a human student), and a mentor agent interacting with students in a multiparty serious game. All of these systems automatically analyze the language and discourse of the students as they interact with the learning environments. The missions of science, technology, engineering, and mathematics (STEM) learning are likely to fail without the integration of SCEM.

Keywords: tutoring, conversational agents, discourse processing, computational linguistics, intelligent tutoring systems, emotions

The two questions that a social psychologist interested in language would want to ask about this essay are “why tutoring?” and “why computers?” Some of the answers to these two questions are obvious. Tutoring is an excellent way to help students learn various skills and subject matters so the students function as good citizens in a society, are financially secure, are healthy and happy, and teach their children well. Computers help people with either repetitive or exceedingly complex tasks that humans try to avoid. So, what could be better? A computer tutor would result in less poverty and more satisfaction among citizens, families, and the workforce.

However, it is the counterintuitive answers that are no doubt most illuminating to readers of this Handbook. Tutoring intersects with social, cognitive, emotional, and motivational (SCEM) mechanisms in unsuspected ways that make the phenomenon worthy of study. The claim could be made that every speech act of a good tutor has tentacles to SCEM processes. Human tutors frequently do not generate intelligent speech acts, and neither

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

do computers, so we can explore the various ways that conversations succeed or fail during tutoring. The practical goal is to build SCEM-sensitive tutors, whereas the theoretical goal is to understand tutoring as a social communication process that helps students learn. Some illustrations of this are foreshadowed here:

1. *The common ground (shared knowledge) between tutor and student is negligible to modest, so there is a struggle to understand each other.* Human tutors often do not realize this, so they yammer on and on, believing mistakenly that they are understood by the student, much like a citizen giving directions to a lost visitor in (p. 492) a city. The student and tutor need to struggle to achieve common ground because there is a large gulf between what the student knows and the tutor knows (Chi, Siler, & Jeong, 2004; Graesser, D’Mello, & Person, 2009; Graesser, Person, & Magliano, 1995). This gap in common ground between computer and student makes it more feasible to build a reasonably smooth conversational computer tutor from the perspective of the student because the student does not readily identify discontinuities in coherence.

2. *The roles of the tutor and student are frequently misunderstood, particularly when the tutor is accomplished.* The default assumption in our culture is that the student is inadequate in knowledge and skills, so the tutor teaches these skills by supplying information. But, in truth, a good tutor gets the student to do the talking, rather than acting as an information delivery system (Chi, Siler, Jeong, & Hausmann, 2001; Millis et al., 2011; Schwartz et al., 2009). Bad tutors give lectures whereas good teachers get the student to talk. Bad tutors are know-it-all’s whereas good tutors ask questions to encourage the co-construction of an answer (e.g., “Great question! How can we figure out how to answer it?”). Some parents grumble when schools ask their children to tutor other students, but they fail to realize that teaching is an excellent way to learn. The good news for the computer scientist is that a computer can get a student to do the talking by asking questions, changing topics, and throwing the conversational spotlight on the student. Whenever the computer is uncertain about what the student knows, it can finesse the conversation to get the student to talk and do things.

3. *A good tutor strategically violates pragmatic rules of everyday conversation.* Participants in a normal conversation assume that whatever is said is part of the common ground in the discourse space (Clark, 1996), that people ask questions when they do not understand (Van der Meij, 1994), and that people should be polite and avoid face-threatening acts (Brown & Levinson, 1987). However, there are a number of conditions in which a good tutor should cast aside these assumptions and violate these norms (D’Mello & Graesser, 2012a; Person, Kreuz, Zwaan, & Graesser, 1995; Ogan, Finkelstein, Walker, Carlson, & Cassell, 2012). A good tutor is skeptical of the student’s understanding of anything expressed, accepts the fact that the student is reluctant to ask questions, and confronts the student to uncover his or her conceptions of the material. There is a tradeoff between a tutor being confrontational and being supportive. This makes it easier on the computer system: when the computer

is in doubt, it can safely assume that the learner understands and knows very little and that an impolite, probing question has added value.

4. *A good tutor puts the student in cognitive disequilibrium.* Students are in cognitive disequilibrium when there are obstacles to goals, contradictions, unexpected events, unusual contrasts, and uncertainty in decisions. These events inspire thought and are opportunities for learning even though cognitive disequilibrium triggers confusion, surprise, and sometimes frustration (Baker, D’Mello, Rodrigo, & Graesser, 2010; D’Mello, Lehman, Pekrun, & Graesser, in press; Graesser & D’Mello, 2012). It is not always good for the tutor to be polite and supportive because problem solving and the acquisition of deep knowledge tend to be accompanied by the negative affective states of confusion, frustration, anger, and even rage (Barth & Funke, 2010; D’Mello & Graesser, 2012b; Graesser & D’Mello, 2012). Politeness norms and deep learning may be contradictory. Computer scientists may be reassured by this because the best computer systems are good at presenting difficult challenges, whereas an empathetic computer tutor might not be convincing.

5. *A tutoring interaction has a rich space of options rather than following a tightly scripted delivery.* There is a structure to tutoring, as will be described later in this essay. But there is also a wide degree of flexibility and “optionhood” in what the tutor can say and do. The tutor can abruptly change its mind, introduce a new question, and end the immediate exchange without seriously disrupting the flow of conversation. This is easier for computers to execute: when in doubt, the system can start a new conversational thread and change its “mind”—just as human tutors do.

The point of this prelude is that tutoring is an interesting social phenomenon, that the constraints of tutoring have complex interactions with SCHEM mechanisms, and that many folklore intuitions about tutoring are off the mark. This essay conveys how important it is for social psychologists to play a central role in guiding research on human and computer tutoring in the future.

(p. 493) **Social and Pragmatic Characteristics of Human Tutoring and Ideal Computer Tutoring**

This section reviews the social and pragmatic processes of human and computer tutoring. Researchers in education and discourse processes have conducted detailed analyses of human tutoring on a variety of subject matters, with tutors who vary in expertise and students who vary in age and abilities (Chi et al., 2001; Chi, Roy, & Hausmann, 2008; Evens & Michael, 2005; Graesser & Person, 1994; Graesser et al., 1995; Lehman, D’Mello, Cade, & Person, 2012; Lepper, Drake, & O’Donnell-Johnson, 1997; McArthur, Stasz, & Zmuidzinas, 1990; Person et al., 1995; Shah, Evens, Michael, & Rovick, 2002). These studies track the speech acts, actions, and emotions of the student and tutor as they communicate turn by conversational turn during the course of tutoring. Their detailed discourse and pedagogical analyses have unveiled insights into SCHEM mechanisms.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

A deep analysis of human tutoring was motivated by some meta-analyses that showed learning gains of approximately 0.4 sigma when comparing one-on-one human tutoring with classroom controls or other suitable controls over study time and content (Cohen, Kulik, & Kulik, 1982). A sigma is a measure of the impact of an intervention that compares the difference between two means (experimental condition and control condition) in standard deviation units. The expertise of the tutor does matter, but more evidence is needed to support such a claim according to available research. Collaborative peer tutoring shows an effect size advantage of 0.2–0.9 sigma (Mathes & Fuchs, 1994; Topping, 1996), which is slightly lower than older unskilled human tutors. At the other end of the spectrum, accomplished human tutors have an impact of 0.6–2.0 sigma (Bloom, 1984; VanLehn, 2011; VanLehn et al., 2007), but the sample of accomplished tutors was small and the effect sizes depended very much on the subject matter and targeted skills. These studies showing the effectiveness of human tutoring prompted researchers to identify patterns of language, discourse, pedagogy, and social interaction that might explain why tutoring is so valuable in helping students learn.

The research on human tutoring convinced researchers in the computational sciences to build intelligent tutoring systems to simulate tutoring. These systems help students learn by holding conversations with them in natural language and by implementing tutoring strategies that simulate either humans or ideal pedagogical strategies. Prominent examples of these systems are AutoTutor (Graesser, D’Mello et al., 2012; Graesser et al., 2004; VanLehn et al., 2007), CIRCSIM Tutor (Evens & Michael, 2005; Shah et al., 2002), Coach Mike (Lane, Noren, Auerbach, Birch, & Swartout, 2011), Guru (Olney et al., 2012), iS-TART (McNamara, Boonthum, Levinstein, & Millis, 2007), ITSpoke (Litman & Forbes-Riley, 2006), MetaTutor (Azevedo, Moos, Johnson, & Chauncey, 2010), Operation ARIES! (Millis et al., 2011), PACO (Rickel, Lesh, Rich, Sidner, & Gertner, 2002), Tactical Language and Culture System (Johnson & Valente, 2008), and Why-Atlas (VanLehn et al., 2007). These systems were made possible by landmark advances in the fields of computational linguistics (Jurafsky & Martin, 2008), statistical representations of world knowledge (Landauer, McNamara, Dennis, & Kintsch, 2007), corpus linguistics (Biber, Conrad, & Reppen, 1998), educational data mining (Baker & Yacef, 2009), and interdisciplinary connections between psychology and computer science (McCarthy & Boonthum-Denecke, 2012; McNamara, Graesser, McCarthy, & Cai, in press; Pennebaker, Booth, & Francis, 2007).

This section identifies the tutoring strategies that prevail in normal tutoring and that have been implemented in AutoTutor and a number of other computer tutors that help students learn by holding a conversation in natural language. Whereas some human tutoring strategies have been implemented in computer tutors, computer tutors also sometimes implement ideal strategies not exhibited by human tutors. Whatever strategies are implemented in a computer tutor, it is important to identify their foundations in SCHEM mechanisms. Two fundamental questions are asked in this program of research. First, what discourse patterns occur in normal tutoring and ideal tutoring? Second, do these

conversational patterns help students learn? The primary focus of this essay is on the first question.

The Poverty of Common Ground

Successful communication requires the establishment of a common ground (or shared knowledge) between speaker and listener or writer and reader (Clark, 1996; Holtgraves, 2002; Schober & Brennan, 2003). Common ground is minimal when there is a large gap between the expert tutor's knowledge and the student's knowledge. Tutoring conversations are very different when common ground exists or is achieved versus when it deteriorates. This predicament requires a good tutor to (p. 494) detect and repair misunderstandings, but the vast majority of tutors are unable to do so (Graesser et al., 2009). Graesser, D'Mello, and Person (2009) identified five "illusions" of tutoring that reflect the likely deterioration of common ground and how the tutor might repair such deficits. The DeepTutor computer tutor (Rus, 2010) has been designed to overcome these illusions and thereby improve tutoring of physics.

- 1. Illusion of grounding.** The tutor and student hold the mistaken belief that they have shared knowledge about a word, referent, or idea being discussed in the tutoring session. Such grounding is often illusory. A good tutor tries to gauge the student's level of understanding and to troubleshoot potential breakdowns in communication.
- 2. Illusion of feedback accuracy.** The tutor and student assume that their feedback to each other is accurate during their communication. This often is a mistaken belief. When tutors ask comprehension-gauging questions (e.g., "Do you understand?" "Are you following?"), many students mistakenly say "yes," whereas those students with better comprehension tend to say "no." When the student gives error-ridden or vague answers to a tutor's question, the tutor's short feedback ("good," "maybe," "no") is more likely to be positive than negative. A good tutor should not trust the student's answer to a comprehension-gauging question. Similarly, a student should not trust the short feedback of a polite tutor.
- 3. Illusion of discourse alignment.** The tutor and student often mistakenly assume that they are connected on the discourse function of speech acts. For example, tutors often give indirect hints to nudge the student to think in a particular direction ("What about X?"), but students don't interpret these speech acts as hints. Sometimes students ask a question with a presupposed assertion ("Isn't 17 a prime number?"), but tutors answer the questions ("Yes. 17 is a prime number") without giving students credit for their contribution. A good tutor clarifies the epistemological status of the speech acts expressed in a tutoring session.
- 4. Illusion of student mastery.** Both tutors and students assume that the student has mastered much more knowledge than the student has really mastered. For example, it is assumed that the student has mastered a complex conceptualization if the student can express a couple of words or ideas about it. However, this skimpy student articulation hardly goes the distance in understanding a complex conceptualization.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

A good tutor asks follow-up questions to verify student understanding. A good student does not overestimate what he or she knows.

5. *Illusion of knowledge transfer*. The tutor believes that the student understands whatever the tutor expresses; the student assumes that whatever he or she expresses is understood by the tutor. In fact, the student absorbs surprisingly little of what the tutor says and vice versa. A good tutor assumes that nothing he says is understood unless the student expresses or does something that reflects understanding.

The following example illustrates a large gap in common ground between the tutor and student. This came from a tutoring session on research methods.

TUTOR: Do you know how to get the main effect in a factorial design?

STUDENT: Yeah, and I understand that, like, its what the independent variable itself...like a measurement of it, itself. And, um, I know, like, that looking at these, I know how to get it.

TUTOR: Okay, right. So if you have a 3 by 2 factorial design, you look at the variability among marginal means on one independent variable and divide that by the error term.

STUDENT: I see.

Examples such as this illustrate the complexity of tutoring with respect to developing and monitoring common ground between tutor and student. The student's language is ungrammatical, vague, semantically ill-formed, and incoherent—in sharp contrast with the tutor's tight analytical construction. The student politely expresses "I see" after the tutor's description of a main effect, but the likelihood that the student truly understands is near zero. This student's contribution is representative of students' language that a computer tutor needs to handle if it converses with the student in natural language.

Expectation and Misconception-Tailored Dialogue

The tutoring strategies of human tutors follow a systematic conversational structure that is called *expectation and misconception-tailored dialogue* (p. 495) (EMT dialog; Graesser, D'Mello et al., 2012; Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; VanLehn et al., 2007). Human tutors anticipate particular correct answers (called expectations) and particular misunderstandings (misconceptions) when they ask their students challenging questions (or problems) and trace their reasoning. As the students express their answers, which are distributed over multiple conversational turns, their contributions are compared with expectations and misconceptions. The tutors give feedback to the students' answers with respect to matching the expectations or misconceptions. The short feedback consists of positive, neutral, or negative expressions either in words, intonation, or facial expressions. Interestingly, sometimes the words are politely positive (good) or neu-

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

tral (okay), but the intonation or facial expressions are pedagogically accurate; for example, negative after an erroneous or vague student response (Graesser et al., 1995).

After the short feedback, the tutor tries to lead the student to express expectations (good answers) through multiple dialogue moves, such as *pumps* (“What else?”) *hints*, and *prompts* to get the student to express specific words (Graesser, Conley, & Olney, 2012). When the student fails to answer the question correctly, the tutor contributes information as *assertions*. These pump-hint-prompt-assertion cycles are frequently used in tutoring to extract or cover particular sentence-like expectations. Eventually, all of the expectations are covered, and the exchange is finished for the main question or problem. During this process, the student occasionally asks questions, which are immediately answered by the tutor, and the student expresses misconceptions, which are immediately corrected by the tutor. A more accomplished tutor might try to get the student to answer her own questions or correct her own misconceptions, but such self-regulated activities are extremely rare in actual tutoring.

Expectation and misconception-tailored tutoring can be simulated on a computer, as in the case of AutoTutor and Why/Atlas for introductory computer literacy and Newtonian physics (Graesser et al., 2004; Graesser, D’Mello et al., 2012; VanLehn et al., 2007). The success of the computer’s implementing the EMT dialogue mechanism depends on how well semantic matches can be made between the verbal contributions of the student and the list of expectations and misconceptions. Fortunately, advances in natural language processing research have made major progress in the accuracy of these semantic matches in computers. These semantic match algorithms have included keyword overlap scores, word overlap scores that place higher weight on lower frequency words in the English language, scores that consider the order of words, latent semantic analysis cosine values, regular expressions, and procedures that compute logical entailment (Cai et al., 2011; Graesser & McNamara, 2012; Graesser, Penumatsa, Ventura, Cai, & Hu, 2007; Rus & Graesser, 2006). It is beyond the scope of this essay to describe these semantic match algorithms, but, briefly, in these assessments, a semantic match score (between 0 and 1) is made between the verbal contributions of students and the sentence-like expressions of expectations and misconceptions. The question is how well the computer match scores correspond to the match scores of trained human judges. The performance data are impressive in most of these assessments. Cai et al. (2011) recently reported an algorithm with computer semantic match scores correlated .67 with trained human judges, whereas the trained human judges correlated .69 with each other.

In contrast to EMT dialogue, it is very difficult to implement a host of sophisticated pedagogical strategies that humans also rarely exhibit, such as bona fide Socratic tutoring strategies, modeling-scaffolding-fading, reciprocal teaching, frontier learning, building on prerequisites, or diagnosis/remediation of deep misconceptions (Collins, Brown, & Newman, 1989; Palincsar & Brown, 1984; Rogoff & Gardner, 1984; Sleeman & Brown, 1982). In *Socratic tutoring*, the tutor asks learners illuminating questions that lead them to discover and correct their own misconceptions in an active, self-regulated fashion. In *modeling-scaffolding-fading*, the tutor first models a desired skill, then gets the learners to per-

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

form the skill while the tutor provides feedback and explanation. The tutor eventually fades from the process until the learners perform the skill unaided. In *reciprocal teaching*, the tutor and learner take turns working on problems or performing a skill, as well as giving feedback to each other along the way. Tutors who use *frontier learning* select problems and give guidance in a fashion that slightly extends the boundaries of what the learner already knows or has mastered. Tutors who *build on prerequisites* cover the prerequisite concepts or skills in a session before moving to more complex problems and tasks that require mastery of the prerequisites.

There are important implications to the discovery that EMT dialogue is both common in human tutoring and can be implemented in computer tutors. One practical implication is that these computer tutors can (p. 496) fill a critical need in providing human tutoring when the human tutors are unavailable. Available empirical evidence indeed supports the claim that computer tutors with natural language dialogue yield learning gains comparable to trained human tutors, with effect sizes of 0.6–1.0 (VanLehn, 2011; VanLehn et al., 2007; Olney et al., 2012). A second implication is that these computer tutors may do even better than human tutors if they weave in the more sophisticated ideal tutoring mechanisms just described. That is, the best computer tutor may be a hybrid of human strategies and ideal pedagogical tutoring strategies.

Speech Acts and Other Dialogue Moves

A *speech act* is a category of utterance, clause, or sentence that serves a discourse function. Examples of speech act categories are questions, answers, statements (assertions), requests, commands, short reactions, expressive evaluations, greetings, and meta-comments. The speech act categories adopted in the tutoring research are based on theoretical schemes, with eight to ten categories that also can be reliably coded by trained judges (D'Andrade & Wish, 1985; Graesser, Swamer, & Hu, 1997; Lehman et al., 2012; Olney et al., 2003). These categories can be broken down into several dozen subcategories. Computer tutors need to segment the stream of words from the students' contributions into speech acts and then assign them to one of these categories.

The speech acts produced by the human or computer tutor have a somewhat different classification scheme that is motivated by pedagogical considerations. These speech act categories, or what are often called *dialogue moves*, include the following:

- *Short feedback*. The feedback to the student is either *positive* (“yes,” “very good,” head nod, smile), *negative* (“no,” “not quite,” head shake, pregnant pause, frown), or *neutral* (“uh huh,” “I see.”)
- *Pumps*. The tutor gives nondirective pumps (“What else?” “Tell me more”) to get the student to do the talking.
- *Hints*. The tutor gives hints to get the student to do the talking or doing, but directs the students along some conceptual path. The hints vary from being generic statements or questions (“What about X?” “Why not?”) to speech acts that more directly

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

lead the student to a particular answer. Hints are the ideal scaffolding move to promote active student learning while directing the student to focus on important relevant material.

- *Prompts*. The tutor asks a very leading question to get the student to articulate a particular word or phrase. Sometimes students say very little, so these prompts are needed to get the student to say something.
- *Assertions*. The tutor expresses a fact or state of affairs.

Other categories of tutor dialogue moves are self-explanatory, such as *answers* to student questions, *corrections* of student misconceptions, *summaries*, *mini-lectures*, or off-topic *comments*.

One frequent discourse pattern occurs when the tutor tries to get the student to articulate a sentence, such as the Newtonian law “Force equals mass times acceleration.” The tutor starts with a pump, with the hope that the student articulates the law. If not, the tutor gives a hint, such as “What about force?” or “What does force equal?” If the student doesn’t articulate “mass times acceleration,” then the tutor gives a prompt, such as “Force equals mass times what?” If the student still comes up empty, the tutor simply asserts the law by saying “Force equals mass times acceleration.” This pump → hint → prompt → assertion cycle is common in human tutoring and is implemented in AutoTutor.

The distribution of tutor dialogue moves reflects the extent to which the tutor versus the student is contributing to answering the question (or solving the problem). This is measured in the correlations between students’ knowledge about the topic and the likelihood that tutors express these categories. For example, there is a significant positive correlation between student knowledge and the tutor’s positive feedback, pumps, and hints, but a negative correlation with the tutor’s negative feedback, prompts, assertions, and corrections (Jackson & Graesser, 2006). This finding suggests that a researcher can infer what the student knows from the speech acts of tutors who intelligently interact with the student. The tutors’ speech act distributions reveal the extent to which the information is contributed by the student versus the tutor in the collaborative construction of an answer/solution.

Sometimes students misunderstand the discourse function of particular speech act categories, such as hints. When hints are assertions, the student may view them as mere statements of fact and not realize that the tutor is trying to steer the student to a better answer. When hints are questions (such as “What about X?”) and the student does not see what the tutor is driving at, then the student becomes confused and retorts (“Well what *about* X?”). A good tutor minimizes confusion by preceding the hint with a discourse marker that signals the (p. 497) discourse function, such as “Here’s a hint” or “Let me get you back on track.”

Five-Step Tutoring Frame

The tutor frequently implements a Five-Step Tutoring Frame to solve the problem or answer a difficult question (Graesser, Conley et al., 2012; Graesser & Person, 1994; Graesser et al., 1995). The five steps are identified here:

1. Tutor asks a difficult question (or presents the problem).
2. Student gives an initial answer.
3. Tutor gives immediate, short feedback on the quality of the answer.
4. Tutor and student interact through a multiturn dialogue to improve the answer.
5. Tutor assesses whether the student understands the answer.

Step 4 is particularly important because this is when the student and tutor interact dynamically in the collaborative construction of an answer. Step 5 normally involves a comprehension-gauging question by the tutor, such as “Do you understand?” For a variety of reasons, most students incorrectly say “Yes,” even when they do not understand. For example, they may have poor metacognitive assessments of their own knowledge, or they may be reluctant to disappoint the tutor. The more knowledgeable student tends to say “No, I don’t understand” (Chi, DeLeeuw, Chiu, & LaVancher, 1994; Graesser et al., 1995). It may take knowledge to know what you don’t know (Miyake & Norman, 1979). For these reasons, a good tutor should not rely on comprehension-gauging questions to assess the students’ knowledge. Instead, he should ask follow-up question or give follow-up tasks to further assess the students’ understanding.

The first three of these steps often occur in a classroom context, using easier short-answer questions. The teacher asks a question, a student is called on to give a quick answer, and the teacher gives positive or negative feedback (Sinclair & Coulthard, 1975). Perhaps tutoring is better than classroom teaching because more difficult questions are asked or because of the collaborative interaction in step 4. Research has not yet determined which of these explanations is most plausible.

The Structure of a Tutor Conversational Turn

Human tutors and AutoTutor structure most of their conversational turns in three constituents or *slots*. The first slot is the positive, neutral, negative feedback on the quality of the student’s last turn. The second slot advances the interaction with either prompts for specific information, hints, assertions of expected answers, corrections of misconceptions, or answers to the student’s question. The third slot is a cue to shift the conversational floor from the tutor to the student. For example, the human tutor cues the student with a question, an intonation signal, or a gesture to encourage the students to talk. Without this cue, there is a standoff of silence, as each waits for the other to speak. In summary, the structure of many tutor turns follows this computational rule:

Tutor Turn → Short Feedback + Dialogue Advancer + Floor Shift

Two examples of a tutor's conversational turn are below:

- "That's right. 17 is a prime number. Why is that important?"
- "Not quite. The difference between acceleration and velocity is [hand gesture]."

The short feedback that begins most tutor conversational turns has been investigated in considerable detail (Evens & Michael 2005; Fox, 1993; Graesser et al., 1995; Person et al., 1995). Humans are prone to give positive feedback after low-quality or error-ridden student turns, perhaps to be polite or not discourage the student from contributing. Experienced human tutors are allegedly more discriminating and less concerned with politeness (Lehman et al., 2012), but systematic comparisons of tutors with varying expertise have not been conducted on short feedback. Graesser, D'Mello, and Person (2009) discussed the plausible hypothesis that students want their tutor feedback to be accurate, with politeness taking a back seat. AutoTutor sometimes does not accurately interpret student contributions, so it occasionally gives incorrect short feedback. Students sometimes become annoyed when this occurs, occasionally dismissing the utility of AutoTutor altogether (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008). Another observation from these sessions with AutoTutor is that students want decisive feedback (*yes* vs. *no*) rather than evasive or indecisive feedback (*possibly*, *uh huh*, *okay*). A polite or wishy-washy computer tutor does not seem to be as desirable as a decisive one. We speculate that the students' assumptions about pragmatic ground rules and communication may be very different for human as opposed to computer tutors.

(p. 498) **Role of Motivation and Emotions**

In addition to cognition, emotions and motivation play a critical role in the learning process. There is a rich literature on the role of motivational and attitudinal traits on learning. Enduring traits that are relevant to learning have tapped constructs of motivation, self-concept, and goal orientation (Daniels et al., 2009; Deci & Ryan, 2002; Dweck, 2002; Frenzel, Pekrun, & Goetz, 2007; Linnenbrink, 2007; Pekrun, Elliot, & Maier, 2006; Schutz & Pekrun, 2007). For example, motivational theorists often contrast mastery-oriented learners who want to understand the material versus performance-oriented students who want good grades. Mastery-oriented learners are more persistent, even in the case of failure, whereas performance-oriented learners have more negative emotions after failures and tend to avoid such challenges. Self-concept theorists contrast students who view themselves as having a trait of being good or bad at learning a subject matter versus those who believe that effort will pay off at learning virtually any subject matter; the former tend to get bored with learning activities much sooner than the latter. Another contrast that has been made is between adventuresome learners, who want to be challenged with difficult tasks, take the risks of failure, and manage negative emotions, and cautious learners, who tackle easier tasks, take fewer risks, and minimize failure and the resulting negative emotions (Clifford, 1991; Meyer & Turner, 2006). An ideal tutor should

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

be sensitive to these psychological learner traits but, interestingly, this is an understudied research area.

Lepper was among the first to directly investigate the role of motivation and emotions in tutoring (Lepper et al., 1997; Lepper & Woolverton, 2002). An INSPIRE model was proposed to promote this integration. This model encouraged the tutor to nurture students by being empathetic and attentive to students' needs, to assign tasks that are neither too easy nor too difficult, to give indirect feedback on erroneous student contributions rather than harsh feedback, to encourage the students to work hard and face challenges, to empower the students with useful skills, to give the students choices, and to pursue topics that the students are curious about. One interesting tutor strategy is to assign an easy problem to the student, but claim that the problem is difficult and encourage the student to give it a try anyway. When the student readily solves the problem, she builds self-confidence and self-efficacy in conquering difficult material.

Research on intelligent tutoring has recently concentrated on the moment-to-moment emotions that students experience during learning (Baker, D'Mello, Rodrigo, & Graesser, 2010; Calvo & D'Mello, 2010; Conati & Macclaren, 2010; Du Boulay et al., 2011; D'Mello & Graesser, 2010; D'Mello, Craig, & Graesser, 2009; Graesser & D'Mello, 2012; McQuiggan, Robison, & Lester, 2010; Picard, 2010; Woolf et al., 2009). The emotional states are polled every few seconds by different methods of measurement, such as behavioral observations, learner self-reports, expert judges, physiological recordings, facial expressions, body movements, and language/discourse of the tutor-student interaction. This research has revealed that a small set of *learning-centered* emotions dominate emotional experiences during learning: confusion, flow (engagement), boredom, and frustration, with delight and surprise occurring less frequently and anxiety occurring under testing contexts. These learner-centered emotions are very different from the six *basic* emotions investigated by Ekman (1992), which are readily manifested in facial expressions: sadness, happiness, anger, fear, disgust, and surprise.

Interesting, investigations with AutoTutor have revealed that the language and discourse of student and tutor are very diagnostic of the emotional states of the student (D'Mello et al., 2008; D'Mello & Graesser, 2010, 2012c; Graesser & D'Mello, 2012). In fact, when all of the learning-centered emotions are considered, language/discourse features predict learner emotional states as well as facial expressions and better than body posture. The positive emotion of *flow* tends to occur when the learner is quickly producing information during his or her conversational turns and receives positive feedback from the tutor. The negative emotion of boredom tends to occur later in the tutoring session, usually when AutoTutor is presenting information (asserting, summarizing, lecturing); presumably, by then, the student is getting bored from information overload and lack of activity. Frustration tends to occur when students are producing information that they believe is on the mark, but for which they receive negative feedback because AutoTutor does not give the student due credit. Confusion tends to occur when discourse cohesion is low (e.g., AutoTutor uses a term or pronoun the student does not understand), the learner does not produce much information, the student is slow to respond, the feedback is negative or con-

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

tradictory, and the student does not understand AutoTutor's hints. At these moments, (p. 499) the student is deep in thought and experiencing cognitive disequilibrium. As mentioned earlier, confusion has been the highest predictor of learning gains from AutoTutor compared with the other learning-centered emotions (D'Mello et al., in press; Graesser & D'Mello, 2012). Confusion occurs when the learner experiences cognitive disequilibrium, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, and uncertainty. Of course, persistent, hopeless confusion presumably has little pedagogical value.

The detection of student emotions is an important advance, but it is not particularly useful unless the tutor can intelligently respond to student emotions. Therefore, an emotion-sensitive version of AutoTutor, called Affective AutoTutor, was developed to automatically detect student emotions based on multiple communication channels and to respond to students' emotions by selecting appropriate discourse moves and displaying emotions through facial expressions and speech (D'Mello & Graesser, 2012a, 2012b; D'Mello, Craig, Fike, & Graesser, 2009). Student emotions are automatically tracked by facial expressions, body posture, language, and discourse interaction (D'Mello & Graesser, 2010). The primary student emotions that Affective AutoTutor tries to handle strategically are confusion, frustration, and boredom because these are the emotions that run the risk of leading to disengagement from the task. The tutor continues business as usual when the student is emotionally neutral, in the state of flow/engagement, or is experiencing the fleeting emotions of delight and surprise; there is no need to respond to these affective states in any special way.

Affective AutoTutor uses a complex set of algorithms that determine how it responds to student emotions. It is beyond the scope of this essay to cover these mechanisms, but a few examples illustrate how this can be accomplished. When frustration is detected, the tutor expresses supportive empathetic comments ("The material is difficult but I believe you can get it") to enhance motivation, in addition to supplying hints or prompts to help the student contribute to answering the question. Frustration needs to be handled to prevent the student from transitioning to boredom and disengagement in a downward spiral. When the student is bored, the tutor responds by giving more engaging material (some razzle dazzle) or challenging the student with more difficult material (for high-knowledge students) or easier material (for low-knowledge students). Regarding confusion, the cognitive disequilibrium framework predicts that confusion is a critical juncture in the learning process, one that is sensitive to individual differences. Some students may give up when experiencing confusion because they perceive themselves as not being good with the subject matter or to avoid negative feedback. For these kinds of students, encouragement, hints, and prompts are allegedly the best strategy for helping them get over the hurdle. Other students treat confusion as a challenge to conquer and expend cognitive effort to restore equilibrium; these students need no special treatment. Affective AutoTutor discriminates between these two types of students by the automated detection of confusion together with the quantity and quality of student contributions in the tutorial interaction.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Affective AutoTutor has an emotion generator that displays a limited set of emotions when the system issues a launch command. With the emotion generator activated, the agent speaks with an intonation that is properly integrated with facial expressions that display particular emotions. The agent nods enthusiastically with approval and expresses positive feedback language after the student makes a correct contribution. The agent shakes its head or has a skeptical facial expression when student contribution is of low quality. There is an empathetic verbal message, kind facial expressions, and an encouraging demeanor when the student needs support. A small set of emotion displays like these go a long way toward conveying the tutor's emotions.

A study was conducted to test the impact of different versions of Affective AutoTutor on learning gains (D'Mello & Graesser, 2012a). The study compared the original AutoTutor without emotion tracking and emotional displays to the emotionally supportive Affective AutoTutor version. The supportive Affective AutoTutor used polite and encouraging positive feedback ("You're doing extremely well") or negative feedback after a low-quality student contribution ("Not quite, but this is difficult for most students"). When a student offered low-quality contributions, the tutor attributed the problem to the difficulty of the material for most students rather than blaming the student. There was also a *shake-up* version of Affective AutoTutor. This version tried to shake up student emotions by being playfully cheeky and telling the student what emotion he or she was experiencing ("I see that you are frustrated"). The simple substitution of this feedback dramatically changed AutoTutor's personality.

(p. 500) The different AutoTutor versions systematically influenced the learning of computer literacy, but in a way that depended on the phase of tutoring and the student's level of mastery. During early phases of the tutoring session, the supportive Affective AutoTutor had either no impact (for low-knowledge students) or a negative impact (for high-knowledge students) on learning. During a later phase of tutoring, the supportive AutoTutor improved learning, but only for low-knowledge students. These results suggest that supportive emotional displays by AutoTutor may not be beneficial during the early phases of an interaction when the student and agent are "bonding" but that a supportive tutor is appropriate at later phases for students who have low knowledge and encounter difficulties. In essence, there may be an optimal *time for emoting*. The shake-up AutoTutor was never fully tested because an early study indicated that learning gains were the same as for the original AutoTutor. However, most adults have a positive initial impression of the shake-up AutoTutor because the playful shake-up tutor is engaging, at least initially. Perhaps the shake-up tutor can be motivating when boredom starts to emerge for more confident, high-knowledge learners, but this needs to be tested in future research.

Some recent computer tutoring systems have given their agents a variety of personalities that hold some promise in influencing motivation, emotions, and learning. Researchers have explored the advantages and liabilities of a variety of personalities: polite, empathetic, assertive, cool, playful, and even rude (Baylor & Kim, 2005; Ogan et al., 2012; Person, Burke, & Graesser, 2003; Wang et al., 2008). There presumably are times when a polite and supportive agent is best; namely, when students lack confidence and need a boost.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

There are other times when a playfully rude tutor will keep the learner entertained while serious content is smuggled into the material. The delicate balance between play and serious content is a frontier for future research.

Tutoring in Multiparty Conversations

Tutoring often occurs in a group context, so the logical next step is to have computer agents participate in these groups. A human can learn vicariously by observing other agents interacting in ways that model ideal learning processes and collaborative reasoning. A human can interact with a peer during learning or even teach the peer agent while the tutor agent intervenes periodically. A human can be tutored by an AutoTutor agent, with a student agent periodically entering the conversation and illustrating ways that students can interact with the AutoTutor. A human can coach a student agent to help him or her take a test administered by a teacher agent. All of these possibilities have been studied in computer learning environments such as Operation ARIES! (Halpern et al., 2012; Millis et al., 2011), iSTART (McNamara et al., 2007), and Betty's Brain (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010; Schwartz et al., 2009). Larger group interactions with agents have developed in Tactical Language and Culture Training System (Johnson & Valente, 2008), Crystal Island (Rowe, Shores, Mott, & Lester, 2010), River City (Ketelhut, Dede, Clarke, Nelson, & Bowman, 2007), EcoMUVE (Metcalf, Kamarainen, Tutwiler, Grozner, & Dede 2011), and other systems (Kim et al., 2009; Lane et al., 2011). The group helps the student learn how to learn in a social world.

The following section describes two learning environments with multiple agents interacting with a human. Operation ARIES! incorporates *trialogs* that have a human agent holding a conversation with a student peer agent and a tutor agent (Figure 1). AutoMentor is a new system under development that has a computer mentor interacting with a group of three to five students playing a simulation game on urban sciences. The use of agents in group conversations can add considerable sophistication to the learning environment and, in some ways, be easier to implement than one-on-one tutoring.

Conversational Trialogs with Operation ARIES!: Two Agents Are Better than One

Operation ARIES! (Halpern et al., 2012; Millis et al., 2011) teaches scientific critical thinking through a series of game modules with two or more animated pedagogical agents. In each of the modules, three-way conversations occur between the human, a student agent (named "Glass"), and a tutor agent (named "Dr. Quinn"). ARIES is an acronym for Acquiring Research Investigative and Evaluative Skills; it was subsequently renamed Operation ARA when it became commercialized by Pearson Education as a serious game. It takes approximately 20 hours to complete ARIES and 7 hours to complete ARA.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems



Fig. 30.1 Case study module in operation ARIES! with three-way conversations among the human and a tutor agent (named "Dr. Quinn,") as well as two new agents joining Dr. Quinn: a fellow student agent (named "Tracy") an alien defector (named "Broth"). The human player and Tracy compete against each other for the honor of going forward to the next module, where they will interrogate aliens. Broth wants peace with Human "Bean" and is observing the sessions.

ARIES has three major phases: training, case study, and interrogation. The training phase consists of an e-book, multiple-choice questions, and tutorial dialogues that teach scientific concepts involved in psychology, biology, and chemistry. Students receive training on twenty-one core scientific concepts (p. 501) while completing this phase and subsequent phases in the serious game. Example core concepts are hypotheses, operational definitions, independent variables, dependent variables, random assignment, subject bias, and correlation versus causation. In the case studies phase, players apply what they learned in the training phase to realistic examples of flawed research. Specifically, they critique research cases in the media (news reports, blogs, television) that exhibit bad science by identifying which of the twenty-one core concepts are violated. For example, one case study described an experiment that tested a new pill that purportedly helps people lose weight, but with no control group (Figure 30.1). In the interrogation phase, players learn how to ask scientists pointed questions about their research to inquire whether they are violating the core concepts. The storyline is advanced by videos, news flashes, and e-mails that are interspersed among the learning activities to build suspense, surprise, and curiosity in an emerging narrative with a plot. It begins with the player joining the Federal Bureau of Science (FBS) as an agent-in-training and concludes with the player helping save the world from aliens who are trying to take over the world by spreading bad science (Millis et al., 2011). Empirical studies have shown that ARIES helps students learn research methods (Forsyth et al., 2012; Halpern et al., 2012; Millis et al., 2011).

Three types of dialogues are implemented in ARIES and launched under specific conditions. They are (1) vicarious learning (learning by observing agents), (2) tutorial learning (learning by being tutored by a tutor agent), and (3) learning through teaching (learning by teaching a fellow student agent). The type of dialogue that occurs for a particular core con-

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

cept is based on the level of knowledge exhibited by the player earlier in the game. Low knowledge triggers vicarious learning trialogs, intermediate knowledge triggers tutorial learning, and high knowledge triggers learning through teaching.

There is a research precedent for assigning these trialog categories to students on the basis of their knowledge. Observational learning of tutorial sessions enhances learning, particularly for low prior knowledge students when they watch a tutorial conversation and comment on what is being learned rather than participating directly (Chi et al., 2008; Craig, Chi, & VanLehn, 2009; Craig, Sullins, Witherspoon, & Gholson, 2006). In the ARIES trialogs with vicarious learning, human students do participate but only minimally: at one or (p. 502) two times during the interaction between the student agent and tutor agent, the tutor turns to the human student and asks what he or she thinks by asking a simple question (e.g., yes-no answer or a choice among three options). Drawing the human in through the use of a question ensures that the human student stays connected and engaged. Nevertheless, high-knowledge students do not benefit as much from watching a tutorial dialog with minimal involvement. They need to be more active by generating rather than observing information. Indeed, “playing teacher” is particularly effective for high-knowledge students, which justifies their being assigned the learning-through-teaching trialogs. Some learning environments contain “teachable agents” that require the human student to teach a computerized agent. For example, Betty’s Brain (Biswas, Leelawong, Schwartz, & Vye, 2005; Biswas et al., 2010; Schwartz et al., 2009) requires the human student to teach the Betty agent about causal relationships in a biological system that are depicted in a conceptual graph. Betty has to get the conceptual graph correct in order to pass a multiple-choice test. When she fails, the human interacts with Betty to improve her conceptual graph and thereby improve her scores. Another mentor agent guides this interaction through hints and suggestions. The learning-through-teach trialogs are similar to these teachable agents in many ways. The human student teaches the student agent, with the tutor agent periodically entering the conversation to guide a productive interaction. It takes a sufficient amount of knowledge to produce the content to teach the student agent, so only high-knowledge students receive such trialogs.

ARIES incorporates the EMT dialog mechanism just as the AutoTutor does. That is, the goal of the trialog is to help the player articulate a specific expectation (sentence) in the exchange, such as “A scientific hypothesis must have a prediction that can be tested.” The accuracy of matching the students’ contributions in natural language with the expectation is quite high (Cai et al., 2011), as discussed earlier. The trialog begins with a question (e.g., “What is a hypothesis?”). If the human student answers it correctly, then the trialog quickly finishes and the student gets full credit (100 percent), indicating that this core concept is mastered for the unit. However, if the student’s answer is off, then the tutor agent gives a hint, such as “What about testing a hypothesis?” If the student answers correctly (“A prediction is tested when there is a hypothesis”) then the student gets partial credit (67 percent). Otherwise, the tutor agent gives a leading prompt question, such as “What is tested when there is a scientific hypothesis?,” with the hope that the student fills in the word “prediction” and thereby gets some credit (33 percent). If the human student is incorrect (receiving 0 percent credit), then the student agent can barge in and

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

give the correct answer, after which the tutor agent gives positive feedback to the student agent. Instead of giving negative feedback to the incorrect human student, the tutor gives positive feedback to the correct student agent. This promotes politeness and avoids face-threatening negative feedback to the human. Thus, *two agents are better than one*.

The trialogs can be arranged to press the envelope of learning and social interaction even further, as shown here.

- *Requests for a summary*. After the trialog covers the expectation, the tutor or student agent can request a summary from the human student, such as “Could you summarize what a hypothesis is?” This encourages the student to stay engaged and generate information, in addition to providing a repetition of the learning experience. Attention, generation, and repetition are strongly linked to learning.
- *Requests for verification*. The tutor or student agent can request the human student to verify whether a statement is true or false, such as “Do hypotheses require a prediction?” This is another way to assess the student’s knowledge, keep the student engaged, and provide a repetition. These questions may also be sincere, in the sense that the questioner does not know the answer. Whereas the tutor agent may understand what a hypothesis is and believe the human agent does also, that does not mean that the student agent understands what a hypothesis is. Again, two agents are better than one.
- *Student agent barging*. At various points during the trialog, the student agent can barge in and interrupt the thread of exchange between the human and tutor agent with a question or other speech act. Similarly, the tutor agent can barge into the exchange between the human and student agent. This is convenient when the main exchange between two parties is deteriorating. The other agent can interrupt and say, “Wait a minute, what is being tested?”
- *Student agent echoing*. There may be some uncertainty to whether the human student’s contribution has a sufficiently high match to a particular expectation (or misconception). For example, the human student might say “A hypothesis predicts a test outcome,” which (p. 503) does not perfectly match the expectation “A hypothesis has a prediction that can be tested.” The computer is uncertain about what the human means, so the student agent can echo what the computer thinks the human meant by asking “Did you mean that a hypothesis has a prediction that can be tested?” and waiting for the human to answer. The tutor agent then gives feedback (positive or negative) to the students if the human answers “yes” and neutral feedback (“okay”) if the human answers “no.”
- *Requests for clarification*. Another approach to handling uncertainty in what the human student is saying is for one of the agents to request clarification, such as “I don’t understand,” “Could you rephrase that?,” “Could you be more precise?” Indeed, agents are allowed to persistently say they don’t understand, just as people do: “I still don’t understand but let’s move on.”

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

- *Assigning credit and blame.* One possible scheme praises the human and blames the student agent. The tutor can give positive feedback whenever human contributions match expectations. When the human expresses something that matches a misconception, the student agent can partially echo that and the tutor then gives negative feedback to the student agent rather than to the human. Consider the exchange below:

TUTOR: What is a hypothesis?

HUMAN: A claim that can't be tested but predicted.

STUDENT AGENT: A hypothesis cannot be tested.

TUTOR: That is incorrect. A hypothesis can be tested.

HUMAN: But a hypothesis has a prediction.

TUTOR: Right. A hypothesis has a prediction.

In this fashion, the student agent gets the brunt of negative attributions, whereas the human is blessed with positive feedback. This is also a way of decomposing complex utterances that humans express. The student agent can echo pieces of the complex human statement, and the plausibility of each piece can be evaluated by the tutor.

- *Staging cognitive disequilibrium.* Earlier, we discussed the impact of cognitive disequilibrium on confusion, thought, and learning gains. In ARIES, cognitive disequilibrium is planted by manipulating whether or not the tutor agent and the student agent contradict each other during the trialog or express claims that are incorrect (D'Mello et al., in press; Lehman et al., 2011). In the case studies, the tutor agent and student agent deliberate with the student on (a) whether there was a flaw in the study and (b) if there is a flaw, which aspect of the study was flawed. The tutor agent expresses a correct assertion, and the student agent agrees in the *True-True* control condition, whereas the two agents agree on an incorrect assertion in the *False-False* condition. The tutor expresses a correct assertion, and the student agent disagrees with an incorrect assertion in the *True-False* condition, whereas the opposite is the case in the *False-True* condition. As predicted, the contradictory conditions create cognitive disequilibrium, confusion, and often better learning than the control True-True condition.

These studies of ARIES illustrate ways in which trialogs can enhance the learning experience and richness of the social interaction. They show how two agents can be better than one. Other learning technologies have also shown some powerful advantages of multiple agents. As discussed earlier, Betty's Brain has a teachable agent in a learning-by-teaching environment where (a) students teach a computer agent (Betty) to construct a causal map through a visual representation of knowledge, and (b) a teacher agent (Mr. Davis) evaluates Betty's learning performance by asking the agent to answer questions and take quizzes. Betty's Brain has prompts in the conversations between Betty, the teachable agent, and Mr. Davis. The goal is to acquire a deeper mental model, better question-asking skills, and enhanced self-regulated learning (Biswas et al., 2010; Leelawong & Biswas, 2008; Schwartz et al., 2009). Another system called iSTART (Interactive Strategy Training for Active Reading and Thinking) is designed to help students learn metacomprehension strategies that facilitate deeper comprehension in reading (McNamara et al., 2007). A teacher agent and student peer agent interact with the human using natural language to

model, provide feedback, and discuss reading strategies that improve comprehension of difficult science texts. As in the case of ARIES, both Betty's Brain and iSTART have shown learning gains in a series of studies tested on thousands of students. However, this essay focuses on language and dialogue rather than learning gains.

AutoMentor: Planting a Computer Agent in a Multiparty Serious Game

Multiparty games and simulations have become more popular in recent years, so some researchers have attempted to develop serious games that build (p. 504) on these social media. In some systems, a single student experiences a virtual world with several computer agents that change over the course of several hours. Examples of serious games with a single human interacting with multiple agents are Tactical Language and Culture Training System (Johnson & Valente, 2008), Crystal Island (Rowe et al., 2010), and Coach Mike (Lane et al., 2011). In other cases, students communicate on the web in a chat room or some other form of computer-mediated communication as they interact in virtual worlds with complex simulations. Notable examples of these are River City (Ketelhut et al., 2007), EcoMUVE (Metcalf et al., 2011), and Urban Science (Shaffer, 2006).

Shaffer (2006) has argued that student learning is severely limited in simulation game environments if there is no mentorship and expertise provided from professional stakeholders that guide students as they decide what to do and justify their decisions (Hatfield & Schaffer, 2010; Shaffer, 2006). Shaffer's epistemic games group has therefore collected data with human mentors on some of his multiparty games (Urban Science, Land Science). The game helps students understand the kinds of problems and problem solving that socially valued professions routinely engage in. This includes how the development of cities and suburbs are influenced by zoning, roads, parks, housing, and economic investment. It includes findings in science that can be communicated in justifications of decisions. Land Science was designed to simulate a regional planning practicum experience for students. During the 10-hour game, students play the role of interns at a fictitious regional planning firm (called Regional Design), where they make land use decisions to meet the desires of virtual stakeholders who are represented by nonplayer characters (NPCs). Students are split into groups and progress through a total of fifteen stages of the game in which they complete a variety of activities including a virtual site visit of the community of interest, where they explore the history of the site, the ecology of the area, and the desires of different stakeholder groups. The students get feedback from the stakeholders and use a custom designed geographic information system (iPlan) to create a regional design plan. Throughout the game, players communicate with other members on their planning team, as well as with a mentor (i.e., an adult who represents a professional planner with the fictitious planning firm) through the use of a chat feature embedded in the game.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

We have recently collaborated with Shaffer's group to build a computer mentor called AutoMentor (Shaffer & Graesser, 2010). AutoMentor performs automated language and discourse analyses on groups of 3–5 students who interact with each other through computer-mediated communication (chat) as they work on the simulation. AutoMentor periodically makes suggestions on what the human mentor can say next to facilitate student collaboration; the human decides whether to accept these suggestions. The eventual goal, after multiple rounds of development and testing, is to have the AutoMentor replace human mentors. It should be noted that AutoMentor is still in progress and has not been tested empirically. Our aim here is to discuss some of the significant challenges that need to be addressed in developing AutoMentor.

The first challenge was the handling of chat language (called *chatese*), which is replete with acronyms, emoticons, telegraphic expressions, nicknames, figurative language, spelling errors, ungrammatical expressions, and other deviations from standard English. We were fortunate to learn that the rate of chatese is lower in group discussions than in one-on-one chats, that conversational turns are short (about eight words), and that there was no need to use our processing-intensive syntactic parsers on the language analysis. Word spotting, word sequences, and statistical representations of semantics (such as latent semantic analysis) were adequate for performing the analyses of individual student sentences and conversational turns. Conventional data mining procedures (Baker & Yacef, 2009; Witten, Frank, & Hall, 2011) were adequate to classify student sentences into speech act categories (Moldova, Rus, & Graesser, 2011), personality of participants (Keshtkar, Burkett, Li, & Graesser, 2012), and the epistemic categories of epistemic games (Hatfield & Shaffer, 2010). This is an important engineering feat, but the task remains to align these results with theoretical claims.

The second challenge was detecting conversational patterns among students as they interact. We needed to have a good discovery methodology to do this. We turned to *state transition networks* (STNs) to specify the conversational transitions both *within* and *between* speech participants with respect to who speaks and the associated speech act categories. Adjacent speech acts between speakers are known to follow conversational constraints documented decades ago (Sacks, Schegloff, & Jefferson, 1974). Discourse acts in educational contexts have been documented in great detail in the context of classroom discourse (Gee, 1999; Nystrand, 2006; Sinclair & Coulthard, 1975) and human tutoring (Cade, Copeland, Person, & D'Mello, 2008; (p. 505) Graesser, D'Mello, & Cade, 2011; Graesser & Person, 1994; Graesser et al., 1997). For example, examples of adjacency pairs in general conversation are (a) Question → Answer, (b) Promise → Acknowledgment, and (c) Expressive Evaluation → Short response. In classrooms, a common sequence is Teacher Question → Student Answer → Teacher Short Feedback, as discussed earlier. (See also the discussion of the five-step tutoring frame.)

The sequences of speech act categories by different speakers were explored to identify frequent conversation patterns in the multiparty interaction. More specifically, the computer classified sentences in the chat interactions into speech act categories and analyzed sequences of the speech acts by various speakers. We identified those sequences

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

that frequently occur over and above the base rate likelihood of speech acts and sequences. One pattern that frequently occurred was within a mentor turn and was very similar to human tutoring:

Mentor Turn—{~~Short Response~~ or Expressive
Evaluation}+(Statement)
+{Question or Request}

Basically, the mentor gives feedback, makes some justifications through statements/assertions, and then asks questions or makes requests to get the students to do something. Another pattern that frequently occurred was the Mentor macromanaging the students' responses. Students ask "What do we do next?" and the tutor gives directives. We had hoped to see other patterns reflective of students having an intellectual discussion, with student sequences of statements, expressive evaluations, and questions, but that was not statistically detectable. We had hoped that the mentor had an "uptake" of student activities by responding to speech acts and actions initiated by students (Nystrand, 2006) and to continue on their line of thinking, but that was also not apparent from our analyses. It appears that considerable scaffolding is needed before intelligent mentors interact with students in multiparty conversations. It does not come naturally.

The third challenge was to figure out when and how AutoMentor would respond when the ideal moment arose. The "when" is nontrivial. When does the mentor intervene with a turn in the sea of student turns in chat? Should it be during a pause of nonactivity, when the students get into an argument, when the students are floundering, and so on? There is no systematic research on these conversational dynamics. The "what" is also crucial with respect to what the mentor expresses. Should students be encouraged with positive feedback, challenged with adversarial remarks (devil's advocate), or given mini-lectures? Again, there is no systematic research base and theory on learning environments with respect to how the mentor should respond. These are questions for future research.

Conclusion

It should be apparent from this essay that SCEM mechanisms drive deep learning and that language and discourse have a central role in these activities. The hard sciences have declared an urgent need for citizens to pursue science, technology, engineering, and mathematics (STEM) educational pathways. We would like to make an equal plea for SCEM in the development of our learning environments. Without SCEM, the STEM mission will die. It is time for a SCEM-STEM alliance.

Acknowledgments

The research was supported by the National Science Foundation (0325428, 633918, 0834847, 0918409, 1108845) and the Institute of Education Sciences (R305A080594, R305G020018, R305C120001, R305A130030). Any opinions, findings, and conclusions or

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding sources.

References

- Azevedo, R., Moos, D., Johnson, A., & Chauncey, A. (2010). Measuring cognitive and metacognitive regulatory processes used during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*, 210-223.
- Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*, 223-241.
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*, 3-17.
- Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion, 24*(7), 1259-1268.
- Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*, 5-115.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biswas, G., Jeong, H., Kinnebrew, J., Sulcer, B., & Roscoe, R. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning, 5*(2), 123-152.
- (p. 506)** Biswas, G., Leelawong, K., Schwartz, D. L., & Vye, N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*, 363-392.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 4-16.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Cade, W., Copeland, J. Person, N. K., & D'Mello, S. K. (2008). Dialogue modes in expert tutoring. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems* (pp. 470-479). Berlin, Heidelberg: Springer-Verlag.
- Cai, Z., Graesser, A. C., Forsyth, C., Burkett, C., Millis, K., Wallace, P., et al. (2011). Tri-olog in ARIES: User input assessment in an intelligent tutoring system. In W. Chen & S. Li

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

(Eds.), *Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems* (pp. 429–433). Guangzhou: IEEE Press.

Calvo, R. A., & D'Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*, 18–37.

Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477.

Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science, 32*(2), 301–341.

Chi, M. T. H., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction, 22*, 363–387.

Chi, M. T. H., Siler, S. A., Jeong, H., & Hausmann, R. G. M. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–534.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clifford, M. M. (1991). Risk taking: Theoretical, empirical, and educational considerations. *Educational Psychologist, 26*(3/4), 263–297.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237–248.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.) *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.

Conati, C., & Maclaren, H. (2010). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*, 267–303.

Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology, 101*(4), 779–789.

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). Deep-level reasoning questions effect: The role of dialog and deep-level reasoning questions during vicarious learning. *Cognition and Instruction, 24*(4), 563–589.

D'Andrade, R. G., & Wish, M. (1985). Speech act theory in quantitative research on interpersonal behavior. *Discourse Processes, 8*(2), 229–258.

D'Mello, S. K., Craig, S. D., Fike, K., & Graesser, A. C. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In J. Jacko (Ed.), *Hu-*

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

man-computer interaction: Ambient, ubiquitous and intelligent interaction (pp. 595–604). Berlin/Heidelberg: Springer.

D'Mello, S. K., Craig, S. D., & Graesser, A. C. (2009). Multi-method assessment of affective experience and expression during deep learning. *International Journal of Learning Technology*, 4(3/4), 165–187.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B. T., & Graesser, A. C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1–2), 45–80.

D'Mello, S. K., & Graesser, A. C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction*, 20, 147–187.

D'Mello, S. K., & Graesser, A. C. (2012-a). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2, 23:2–23:39.

D'Mello, S. K., & Graesser, A. C. (2012-b). Dynamics of affective states during complex learning. *Learning and Instruction*, 22, 145–157.

D'Mello, S. K., & Graesser, A. C. (2012-c). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5, 304–317.

D'Mello, S. K., Lehman, S., Pekrun, R., & Graesser, A. C. (in press). Confusion can be beneficial for learning. *Learning and Instruction*.

Daniels, L. M., Pekrun, R., Stupnisky, R. H., Haynes, T. L., Perry, R. P., & Newall, N. E. (2009). A longitudinal analysis of achievement goals: From affective antecedents to emotional effects and achievement outcomes. *Journal of Educational Psychology*, 101, 948–963.

Deci, E. L., & Ryan, R. (2002). The paradox of achievement: The harder you push, the worse it gets. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61–87). Orlando, FL: Academic Press.

Du Boulay, B., Avramides, K., Luckin, R., Martinez-Miron, E., Mendez, G. R., & Carr, A. (2011). Towards systems that care: A conceptual framework based on motivation, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 20(3), 197–229.

Dweck, C. S. (2002). The development of ability conceptions. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation: A volume in the educational psychology series* (pp. 57–88). San Diego, CA: Academic Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169–200.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Evens, M. W., & Michael, J. A. (2005). *One-on-one tutoring by humans and computers*. New York: Routledge/Taylor & Francis.

Forsyth, C., Pavlik, P., Graesser, A. C., Cai, Z., Germany, M., Millis, K., et al. (2012). Learning gains for core concepts in a serious game on scientific reasoning. In K. Yacef, O. Zañane, H. HersHKovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 172–175). Chania, Greece: International Educational Data Mining Society.

Fox, B. A. (1993). *Human tutorial dialogue*. New Jersey: Lawrence Erlbaum.

(p. 507) Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction, 17*, 478–493.

Gee, J. P. (1999). *An introduction to discourse analysis theory and method* (2nd ed.). New York: Routledge.

Graesser, A. C., Conley, M. W., & Olney, A. M. (2012). Intelligent tutoring systems. In S. Graham & K. Harris (Eds.), *APA educational psychology handbook: Vol. 3. Applications to Learning and Teaching* (pp. 451–473). Washington, DC: American Psychological Association.

Graesser, A. C., & D'Mello, S. K. (2012). Emotions during the learning of difficult material. In B. Ross (Ed.), *The psychology of learning and motivation*, vol. 57 (pp. 183–225). New York: Elsevier.

Graesser, A. C., D'Mello, S. K., & Cade, W. (2011). Instruction based on tutoring. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 408–426). New York: Routledge.

Graesser, A. C., D'Mello, S. K., Hu, X., Cai, Z., Olney, A. M., & Morgan, B. (2012). AutoTutor. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation and resolution* (pp. 169–187). Hershey, PA: IGI Global.

Graesser, A. C., D'Mello, S. K., & Person, N. K. (2009). Metaknowledge in tutoring. In D. J. Hacker, J. Donlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 361–382). New York: Taylor & Francis.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A. M., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, and Computers, 36*, 180–193.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition, 33*, 1235–1247.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Graesser, A. C., & McNamara, D. S. (2012). Automated analysis of essays and open-ended verbal responses. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, vol. 1: Foundations, planning, measures, and psychometrics* (pp. 307–325). Washington, DC: American Psychological Association.

Graesser, A. C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed-initiative dialogue in natural language. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 243–262). Mahwah, NJ: Lawrence Erlbaum.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9(6), 495–522.

Graesser, A. C., Swamer, S. S., & Hu, X. (1997). Quantitative discourse psychology. *Discourse Processes*, 23(3), 229–263.

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, 7, 93–100.

Hatfield, D., & Shaffer, D. W. (2010). *The epistemography of journalism 335: Complexity in developing journalistic expertise*. Paper presented at the International Conference of the Learning Sciences (ICLS), Chicago, Illinois.

Holtgraves, T. (2002). *Language as social action: Social psychology and language use*. Mahwah, NJ: Erlbaum.

Jackson, G. T., & Graesser, A. C. (2006). Applications of human tutorial dialog in AutoTutor: An intelligent tutoring system. *Revista Signos*, 39, 31–48.

Johnson, W. L., & Valente, A. (2008). Tactical language and culture training systems: using artificial intelligence to teach foreign languages and cultures. In M. H. Goker (Ed.), *Proceedings of the 20th national conference on Innovative applications of artificial intelligence: Vol. 3* (pp. 1632–1639). Menlo Park, CA: AAAI Press.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Keshtkar, F., Burkett, C., Li, H., & Graesser, A. C. (2012). Detecting players personality behavior with any effort of concealment. In A. Gelbukh (Ed.), *Proceedings of 13th International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 503–515). Heidelberg, Germany: Springer.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Ketelhut, D., Dede, C. J., Clarke, J., Nelson, B., & Bowman, C. (2007). Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfbeck, & H. O'Neil. (Eds.), *Assessment of problem solving using simulations* (pp. 37-58). Mahwah: Erlbaum.

Kim, J., Hill, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., et al. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education*, 19, 289-308.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum

Lane, H. C., Noren, D., Auerbach, D., Birch, M., & Swartout, W. (2011). Intelligent tutoring goes to the museum in the big city: A pedagogical agent for informal science education. In G. Biswas, S. Bull, J. Kay & A. Mitrovic (Eds.), *International journal of artificial intelligence in education* (pp. 155-162). Heidelberg: Springer.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching systems: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, 18(3), 181-208.

Lehman, B., D'Mello, S. K., Cade, W., & Person, N. K. (2012). How do they do it? Investigating dialogue moves within dialogue modes in expert human tutoring. In S. Cerri, W. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 557-562). Berlin Heidelberg: Springer-Verlag.

Lehman, B., D'Mello, S. K., Chauncey, A., Gross, M., Dobbins, A., Wallace, P., et al. (2011). Inducing and tracking confusion with contradictions during critical thinking and scientific reasoning. In S. Bull, G. Biswas, J. Kay, & T. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 171-178). Berlin, Heidelberg: Springer.

Lepper, M. R., Drake, M., & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.

(p. 508) Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.

Linnenbrink, E. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation and engagement. In P. Schutz & R. Pekrun (Eds.), *Emotions in education* (pp. 107-124). San Diego, CA: Academic Press.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Litman, D. J., & Forbes-Riley, K. (2006). Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. *Speech Communication, 48*, 559-590.

Mathes, P. G., & Fuchs, L. S. (1994). Peer tutoring in reading for students with mild disabilities: A best evidence synthesis. *School Psychology Review, 23*, 59-80.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7*, 197-244.

McCarthy, P. M., & Boonthum-Denecke, C. (Eds.). (2012). *Applied natural language processing: identification, investigation, and resolution*. Hershey, PA: IGI Global.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & 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* (pp. 227-241). Mahwah, NJ: Erlbaum.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., Cai, Z. (in press). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, MA: Cambridge University Press.

McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society 13*, 40-53.

Metcalf, S. J., Kamarainen, A., Tutwiler M. S., Grotzer, T. A., & Dede, C. J. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations, 3*, 86-90.

Meyer, D. K., & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review, 18*(4), 377-390.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. C., & Halpern, D. F. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp. 169-196). London, UK: Springer-Verlag.

Miyake, N., & Norman, D. A. (1979). To ask a question one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior, 18*, 357-364.

Moldovan, C., Rus, V., & Graesser, A. C. (2011). Automated speech act classification for online chat. *MAICS, 23*-29.

Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English, 40*, 392-412.

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012). Rudeness and rapport: Insults and learning gains in peer tutoring. In S. A. Cerri & B. Clancey (Eds.), *Pro-*

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

ceedings of the 11th International Conference on Intelligent Tutoring Systems. Berlin: Springer-Verlag.

Olney, A. M., D'Mello, S. K., Person, N. K., Cade, W., Hayes, P., Williams, C., et al. (2012). Guru: A computer tutor that models expert human tutors. In S. A. Cerri & B. Clancey (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 256-261). Berlin: Springer-Verlag.

Olney, A. M., Louwerse, M. M., Mathews, E., Marineau, J., Hite-Mitchell, H., & Graesser, A. C. (2003). Utterance classification in AutoTutor. In J. Burstein & C. Leacock (Eds.), *Building educational applications using natural language processing: Proceedings of the Human Language Technology—North American Chapter of the Association for Computational Linguistics Conference 2003 Workshop* (pp. 1-8). Philadelphia: Association for Computational Linguistics.

Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.

Pekrun, R., Elliot, A., & Maier, M. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98, 583-597.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count*. Austin, TX: LIWC.net (www.liwc.net).

Person, N. K., Burke, D. R., & Graesser, A. C. (2003, June). *RudeTutor: A face threatening agent*. Presented at the 13th Annual Meeting of the Society for Text and Discourse. Madrid, Spain.

Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13(2), 161-188.

Picard, R. (2010). Affective computing: From laughter to IEEE. *IEEE Transactions on Affective Computing* 1, 11-17.

Rickel, J., Lesh, N., Rich, C., Sidner, C. L., & Gertner, A. S. (2002). Collaborative discourse theory as a foundation for tutorial dialogue. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems* (pp. 542-551). Berlin: Springer-Verlag.

Rogoff, B., & Gardner, W. P. (1984). Adult guidance of cognitive development. In B. Rogoff and J. Lave (Eds.), *Everyday cognition: Its development in social contexts*. (pp. 95-116). Cambridge: Harvard University Press.

Rowe, J., Shores, L. R., Mott, B., & Lester, J. C. (2010). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, 20, 166-177.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

Rus, V. (2010). DeepTutor: An intelligent tutoring system that promotes deep learning of science topics. Retrieved from University of Memphis, DeepTutor website: <http://deeptutor.memphis.edu/DeepTutor.Overview.V-RusEtAl.May-2012.pdf>.

Rus, V., & Graesser, A. C. (2006). Deeper natural language processing for evaluating student answers in intelligent tutoring systems. In Y. Gil & R. Mooney (eds), *Proceedings of the American Association of Artificial Intelligence* (pp.1495–1500). Menlo Park, CA: AAAI.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse. In A. C. Graesser, M. A. Gernsbacher, & S. Goldman (Eds.), *Handbook of discourse processes* (pp 123–164). Mahwah, NJ: Erlbaum.

Schutz, P., & Pekrun, R. (Eds.). (2007). *Emotion in education*. San Diego, CA: Academic Press.

Schwartz, D. L., Chase, C., Chin, D., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R. D., Jeong, H., & Wagster, J. D. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). London: Routledge Press.

(p. 509) Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Shaffer, D. W., & Graesser, A. C. (2010). *Using a quantitative model of participation in a community of practice to direct automated mentoring in an ill-defined domain. Workshop proceedings at the Intelligent Tutoring Systems Conference on Intelligent Tutoring Technologies for Ill-defined Problems and Ill-defined Domains*. Pittsburgh, PA: ITS.

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

Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse*. Oxford: Oxford University Press.

Sleeman D., & Brown, J. S. (Eds.). (1982). *Intelligent tutoring systems*. Orlando, Florida: Academic Press.

Topping, K. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321–345.

Van der Meij, H. (1994). Student questioning: A componential analysis. *Learning and Individual Differences*, 6, 137–161.

The Role of Natural Language and Discourse Processing in Advanced Tutoring Systems

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A. M., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31, 3–62.

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2), 98–112.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Burlington, MA: Morgan Kaufmann.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology*, 4, 129–164. (p. 510)

Arthur C. Graesser

Arthur C. Graesser, Department of Psychology, Psychology Department, University of Memphis, Memphis, TN

Fazel Keshtkar

Fazel Keshtkar is an Assistant Professor in the Department of Computer Science at the Missouri State University, Southeast.

Haiying Li

Haiying Li, University of Memphis