Tutorial Dialog in Natural Language

Andrew M. Olney[1], Arthur C. Graesser[1], and Natalie K. Person[2]

[1] University of Memphis, Memphis TN 38152, USA,
aolney@memphis.edu,
WWW home page: http://andrewmolney.name
[2] Rhodes College, Memphis TN 38112, USA

Abstract. This chapter reviews our past and ongoing investigations into conversational interaction during human tutoring and our attempts to build intelligent tutoring systems (ITS) to simulate this interaction. We have previously modeled the strategies, actions, and dialogue of novice tutors in an ITS, called AutoTutor, with learning gains comparable to novice tutors. There is evidence, however, that expert human tutors may foster exceptional learning gains beyond those reported for some categories of human tutors. We have undertaken a rigorous, large scale study of expert human tutors and are using these data to create Guru, an expert ITS for high school biology. Based on our analyses, expert human tutoring has several distinctive features which differ from novice human tutoring. These distinctive features have implications for the development of an expert ITS, and we briefly describe how these are being addressed in Guru.

1 Introduction

The empirical evidence that one-to-one human tutoring is extremely effective compared to classroom environments is well known [1–4]. The effectiveness of one-to-one tutoring raises the question of what makes tutoring so powerful. Three different hypotheses, known as the tutor-centered, student-centered, and interaction hypotheses, have been proposed to answer this question [5, 6]. The tutor-centered hypothesis claims that effective tutoring stems primarily from the actions of the tutor, specifically, the tutor's pedagogical moves are tailored to a given student. In contrast, the student-centered hypothesis places the emphasis on the student, highlighting that students are active participants in the construction of their own knowledge rather than being mere information receptacles. Finally, the interaction hypothesis predicts that the effectiveness of tutoring draws from both tutor and student behavior and their coordination with each other. Interaction has also been emphasized in the collaborative learning literature in which some forms of group learning outperform individual learning [7, 8]. Equally related are collaborative forms of instruction that are dialogue-centric, such as reciprocal teaching [9].

Interest in the interaction hypothesis is growing [5, 6]. However, the interaction hypothesis has long-standing roots in the tutoring literature. An early meta-analysis on a large sample of studies compared human-to-human tutoring with classroom environments and suitable comparison conditions [2]. The vast majority of the tutors in these studies were untrained in tutoring skills and had moderate domain knowledge; they were peer tutors, cross-age tutors, or paraprofessionals, but rarely accomplished professionals. These "unaccomplished" human tutors enhanced learning with an effect size of a .4 standard deviation unit (sigma), or approximately a half letter grade. As one might expect, unskilled human tutors are not prone to implement sophisticated tutoring strategies that have been proposed in the fields of education, the learning sciences, and developers of ITSs (Graesser et al., 1995; Graesser, D'Mello, & Person, 2009; Person et al., 1995). Instead the learning gains from unaccomplished human tutors illustrate the power of dialogue-based interaction for learning.

This chapter reviews our work on conversational interaction during human tutoring and our attempts to build intelligent tutoring systems (ITS) to simulate this interaction. To date, the bulk of our research addresses the strategies, actions, and dialogue of novice tutors [4, 10, 11]. We have implemented novice tutoring in an ITS, called AutoTutor, with learning gains comparable to novice tutors [12, 13]. More recently, we have expanded our investigation to highly accomplished expert human tutors [14, 15]. Our shift in emphasis is driven by a desire to understand what makes accomplished expert human tutors produce exceptional learning gains, as has been previously reported [1]. We have undertaken a rigorous, large scale study of accomplished human tutors, and we are using these data to create Guru, an expert ITS for high school biology. In the following sections we further elaborate this contrast between novice and expert, both in terms of human tutoring and the ITS components required to mimic interaction with novice and expert human tutors.

2 Novice human tutoring

Our initial research on conversational interaction during tutoring focused on novice human tutors. While it may seem counterintuitive to focus on novice human tutors when there is evidence that expert human tutors produce superior learning gains, there are several outstanding reasons why it is important to understand novice human tutoring. Perhaps the primary reason to focus on novice human tutoring is that the bulk of human tutors are novice tutors, e.g. paraprofessionals, cross-age tutors, or peers. Expertise is scarce in general, and tutoring is no exception to this rule. A secondary reason to focus on novice human tutors is the plausible assumption that novice human tutors generate a less sophisticated conversational interaction than expert human tutors. By first studying novice tutors, we make use of a divide-and-conquer approach: an understanding of novice human tutoring should make it easier to understand the complexity of expert human tutoring. Finally, the relative abundance of novice human tutors facilitates the creation of larger representative samples than has often been impractical with expert human tutors. We return to this issue when we discuss expert human tutoring in a later section.

Two samples of novice tutoring were collected and analyzed [4, 10]. The first sample consisted of tutoring sessions on undergraduate research methods, the Research Methods Corpus (RMC). The 3 tutors in the RMC were graduate students who had never tutored for research methods, thus they were truly novice tutors in this domain. The 27 students receiving tutoring participated in two 1-hour sessions each with different tutors, for a total of 54 sessions. Each session was recorded; however, due to video quality, only 44 sessions could be transcribed. The second sample of novice tutoring was in the domain of 7th grade algebra, the Algebra Corpus (AC). The 10 tutors who participated were high school students with an average 9 hours of prior experience in tutoring. The 13 students receiving tutoring participated in 1-hour sessions for a total of 22 sessions. Thus in total the RMC and AC consist of 76 hours of tutoring for 40 students and 13 tutors.

Multiple codings schemes have been developed to analyze the RMC and AC along different dimensions including feedback, tutor examples, Gricean Maxims, student errors, and student questions [4, 10, 16, 17]. Tables 1 and 2 present the primary dialogue moves used by students and tutors across these analyses.

Table 1. Novice Student Dialogue Moves

Move Category	Description
Contribution Quality	
Complete	Student provides complete answer to tutor question.
Partial	Student provides partial answer to tutor question.
Vague	Student provides vague answer to tutor question.
Error-ridden	Student provides error-ridden answer to tutor question.
No Answer	Student fails to provide any answer to tutor question.
Asks Question	
Makes request	Student makes request unrelated to the problem/example.
Counter- clarification	Student needs clarification on tutor's previous statement.
Problem-related	Student asks question directly related to the problem/example.
Other	Any question not assigned to one of the other three question categories.
Misconception	Student states his or her own misconception.
Reminding Example	Student comments on a similar example.
Meta-comment	Student comments on own ability or attribute of problem.
Acknowledgement	Student acknowledges tutor's contribution (e.g., Uh-huh).
Gripes	Student complains.
Think aloud	Student thinks aloud.
Nonverbal	Student makes a nonverbal response (e.g., laughs).
Draw	Student draws on board.
Other	Any speech act not assigned to one of the other student categories.

Table 2. Novice Tutor Dialogue Moves

Move Category	Description
Additional example	
Easier	Tutor provides student an easier example than the previous example.
Difficult	Tutor provides a more difficult example than the previous example.
Equal	Tutor provides an example of equal difficulty with the previous example.
Asks question	
Error-repair	Tutor asks question specifically related to student error.
Directed- Activity	Tutor asks question in order to redirect student's activity.
Leading	Tutor asks question to expose student's misconception.
Counter-clarification	Tutor requests clarification of student's previous statement.
Pump	Tutor pumps student for additional information.
Assessment	Tutor assesses student's knowledge about a particular topic.
Global	Tutor globally assesses student's knowledge (e.g., "Do you understand?")
Other	Any question not assigned to one of the other question categories.
Feedback	
Positive	Tutor gives positive feedback to student.
Negative	Tutors gives negative feedback to student.
Neutral	Tutor gives neutral feedback to student.
Immediate	Tutor provides immediate feedback for a student error.
Delayed	Tutor provides delayed feedback for a student error.
Reminding Example	Tutor comments on a similar example.
Specific Component	Tutor focuses on specific component of current problem/example.
General Level	Tutor discusses current example in more general terms.
Hint	Tutor provides the student with a hint.
Splice	Tutor splices in the correct answer.
Elaborates	Tutor elaborates current problem/example.
Answers	Tutor answers student question.
Rearticulates	
Solution	Tutor rearticulates the current problem's solution.
Representation	Tutor rearticulates the problem's representation.
Affective	
Own ability	Tutor comments on his or her own ability.
Student ability	Tutor comments on student's ability.
Problem	Tutor comments on the difficulty of the problem/example.
General	Tutor makes general empathetic comment.
Gripes	Tutor complains.
Directive	Tutor tells the student what to do.
Draw	Tutor draws on the board.
Nonverbal	Tutor makes some type of nonverbal response (e.g., laughs).
Other	Any speech act not assigned to one of the other tutor categories.

Using the coding scheme in Tables 1 and 2, Graesser and Person's analyses of the RMC and AC uncovered three frequent dialogue structures [4, 10, 18]. These same structures have featured prominently in the work of other researchers conducting fine-grained analyses of tutoring [5, 19, 6, 20, 21]. These three dialogue structures are:

1. 5-step Tutoring Frame
2. Expectation and Misconception Tailored (EMT) dialogue
3. Conversational Turn Management (which includes tutor pedagogical modes)

These three structures are multiply embedded: 3 is embedded in 2, which in turn is embedded in 1. There are two common features across all three of these structures. The first is that the tutor, rather than the student, tends to initiate and guide the conversational interaction. The second common feature is that all three of these structures exist at the level of the problem, rather than across larger spans of the tutorial discourse.

2.1 5-Step Tutoring Frame

The 5-Step Tutoring Frame begins once a problem has been introduced. As indicated by the name, the following five steps are enacted in order:

1. TUTOR asks a difficult question or presents a problem.
2. STUDENT gives an initial answer.
3. TUTOR gives short feedback on the quality of the answer.
4. TUTOR and STUDENT have a multi-turn dialogue to improve the answer.
5. TUTOR assesses whether the student understands the correct answer.

This 5-Step Tutoring Frame involves a great deal of conversational interaction. The structure of the 5-Step Tutoring Frame fosters both collaborative discussion and joint action as the tutor works with the student to iteratively construct a better answer.

The 5-Step Tutoring Frame can be better understood by contrasting it with the Initiate-Respond-Feedback (IRF) sequence typically used in classrooms [22]. The first three steps occur in classroom IRF, but the questions are easier short-answer questions. The classroom IRF sequence consists of the teacher initiating a question, the student giving a short-answer response, and the teacher giving a positive or negative feedback of the response. For example, consider the following IRF example for Newtonian physics.

> TEACHER: According to Newton's second law, force equals mass times what?
> STUDENT: acceleration
> TEACHER: Right, mass times acceleration. Or
> STUDENT: velocity
> TEACHER: Wrong, it's not velocity, it is acceleration.

As the above example illustrates, IRF does not facilitate conversational interaction. The 5-Step Tutoring Frame goes beyond IRF by posing more difficult questions that stimulate the collaborative interactions found in step 4.

2.2 Expectation and Misconception Tailored (EMT) dialogue

Novice human tutors maintain a basic representation of the correct answer to a problem (expectations) as well as some misconceptions that may arise. For example, expectations E1 and E2 and misconceptions M1 and M2 are relevant to the example physics problem below.

> PHYSICS QUESTION: If a lightweight car and a massive truck have a head-on collision, upon which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion, and why?
> E1. The magnitudes of the forces exerted by A and B on each other are equal.
> E2. If A exerts a force on B, then B exerts a force on A in the opposite direction.
> M1: A lighter/smaller object exerts no force on a heavier/larger object.
> M2: Heavier objects accelerate faster for the same force than lighter objects

Expectations and misconceptions form a simple domain model which novice tutors use to select dialogue moves. Expectations are akin to the expert model of model tracing tutors, and misconceptions are likewise analogous to buggy libraries in model tracing tutors.

Novice tutors select dialogue moves based on the status of the current problem's expectations and misconceptions. Hints and prompts direct the student to articulate missing content words, phrases, and propositions. For example, a hint for expectation E1 might be "What about the forces exerted by the vehicles on each other?", which would ideally elicit the answer "The magnitudes of the forces are equal." A corresponding prompt to elicit "equal" would be "What are the magnitudes of the forces of the two vehicles on each other?" As the conversational interaction of the tutoring session unfolds, the student articulates the tutor's expectations in piecemeal (as in the examples given) or directly (for a high ability student). Novice tutors also have some awareness of common misconceptions associated with a problem. Thus when a student articulates a misconception, the tutor identifies the misconception and corrects it.

2.3 Conversational Turn Management

The preponderance of conversational interaction is tutor-led. Student led dialogue can occur when students ask questions, but it is well documented that students rarely ask questions, even in tutoring environments [4, 18]. Tutor-led turns usually consist of three steps. The first step gives positive, neutral, or negative feedback on the student's last answer. The second step advances progress through the current problem, based on the expectations and misconceptions the student has already covered. Thus the second step may be instantiated with prompts for specific information, hints, assertions with correct information, or corrections of misconceptions. The third step signals the student that it is their turn to respond, i.e. via a question, rising intonation, or a gesture.

Novice human tutors use the 5-Step Tutoring Frame, EMT dialogue, and conversational turn management to present challenging problems or questions to the student, adaptively scaffold good answers through collaborative interactions, provide feedback when students express erroneous information, and answer occasional student questions. What is absent are sophisticated pedagogical strategies. According to our systematic analyses of the tutoring process [10, 23, 17], novice human tutoring is not characterized by sophisticated tutoring strategies that have been proposed in the fields of education, the learning sciences, and developers of ITS [24]. In particular, novice tutors rarely engage in pedagogical techniques such as bona fide Socratic tutoring strategies [25], modeling-scaffolding-fading [26], Reciprocal Teaching [9], frontier learning [27], building on prerequisites [28], or diagnosis/remediation of deep misconceptions [29]. This is perhaps unsurprising because these strategies are complex and were not discovered for centuries.

3 AutoTutor

AutoTutor simulates a novice human tutor by holding a conversation with the learner in natural language. The pedagogical framework of AutoTutor was inspired by three bodies of theoretical, empirical, and applied research. These include explanation-based constructivist theories of learning [30, 5, 31, 32], intelligent tutoring systems that adaptively respond to student knowledge [33, 34], and empirical research that has documented the collaborative constructive activities that routinely occur during human tutoring [5, 35, 10, 36, 21]. The pedagogical strategies of AutoTutor are modeled on the novice human tutoring strategies described in Section 2, including the 5-Step Tutoring Frame, EMT dialogue, and conversational turn management.

AutoTutor implements the 5-Step Tutoring Frame by presenting a series of challenging questions or problems that require approximately a paragraph of information to answer correctly. An example question in conceptual physics is, "When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation?" Although a perfect answer to this question is approximately 3-7 sentences in length, the initial answers by actual human learners are typically only 1 word to 2 sentences in length. The conversational interaction afforded by tutorial dialogue is particularly helpful when the student's answer is incomplete. AutoTutor uses the 5-Step Tutoring Frame to assist the learner in the evolution of an improved answer by drawing out more of the learner's knowledge that is relevant to the answer. The dialogue between AutoTutor and the learner typically lasts 50-200 turns (i.e., the learner expresses something, then the tutor, then the learner, and so on), which is on par with the interactivity in human tutoring.

AutoTutor uses expectations and misconceptions as an integral part of its domain model, and selects dialogue moves that elicit expectations and address misconceptions. More specifically, the goal of AutoTutor is to elicit the correct answer from the student. Since the correct answer is a paragraph of information, this goal reduces to eliciting each sentence, an expectation, in the correct answer paragraph. In order to elicit each expectation, AutoTutor generates tutorial dialogue moves including pumps, hints, prompts, and assertions:

Pumps. AutoTutor pumps the student for more information during the early stages of answering a particular question (or solving a problem). The pump signals the student to keep talking, for example using positive feedback (e.g., right, yeah, dramatic head nod), neutral back channel feedback (uh-huh, okay, subtle head nod), and explicit requests for more information (What else?, Tell me more). By encouraging the student to say more, pumping helps expose the student's knowledge while giving the student an opportunity to construct knowledge by herself.

Hints. When the student is having problems answering a question or solving a problem, the tutor gives hints by presenting a fact, asking a leading question, or reframing the problem. Hints cue the student to some relevant feature of the problem without revealing the role of that feature in answering the problem.

Prompts. AutoTutor supplies the student with a discourse context and prompts them to fill in a missing word or phrase. Prompting is a scaffolding device for students who are reluctant to supply information. Students are expected to supply more content and more difficult content as they progress in learning the domain knowledge.

Assertions. AutoTutor gives a summary to an expectation. This summary serves the function of succinctly codifying a lengthy, multi-turn, collaborative exchange when an expectation is covered or a problem step is completed.

It is worth noting the continuum of information provided by the tutor in different types of moves. Moves at the beginning of the list, i.e. pumps and hints, provide less information to the student than moves towards the end of the list, i.e. prompts and assertions. By only giving more information when the learner is floundering, AutoTutor promotes active construction of knowledge [10, 5]. Analysis of AutoTutor experiments shows that deeper questions, i.e. pumps and hints, promote more learning than shallow dialogue moves such as prompts and assertions [37].

AutoTutor assesses the student's answers to these dialogue moves using Latent Semantic Analysis (LSA), a vector space method capable of representing world knowledge [38–41]. In LSA, a word is represented by a fixed size vector of real numbers. A sentence or document is also represented by a fixed size vector, made by summing component word vectors. Words, sentences, and documents can all be compared to each other by comparing their vectors. AutoTutor uses LSA to compare the student's answer to the expectations by comparing the LSA vector of the student's answer to the vectors of the expectations. LSA vectors that are identical have a cosine of 1, but AutoTutor uses a lower threshold, e.g. 0.7, to allow the student some flexibility in their answer. In other words, LSA allows student answers with the same meaning, but different wording, to be recognized as correct answers.

AutoTutor uses conversational turn management to maintain a coherent conversational interaction with the student. The primary mechanisms in AutoTutor for conversational turn management are AutoTutor's speech act classifier and dialogue manager. Before AutoTutor responds to a student, the student's utterance is analyzed to determine its speech act [42]. If a student asks an information-seeking question, AutoTutor launches a subdialogue to answer that question. This subdialogue can consist of multiple rounds of clarification, and even recursively nested subdialogues for more detailed questions [43]. If the speech act is an answer or verification question, AutoTutor gives feedback based on the cosine between the student's LSA vector and the current expectation.

In the second and third steps of conversational turn management, AutoTutor selects and delivers a dialogue move. The specific dialogue move is selected based on the current context of the tutoring session, including the problem that the student is on, the current expectation, and the last dialogue move type generated (e.g. hint). AutoTutor loads this context into its state table, an information state [44], and then processes this state table through a dialogue manager [43]. The dialogue manager is defined by a formal language for describing dialogues together with a corresponding interpreter to execute dialogues in this language. This approach has made it much easier to create new tutorial dialogue patterns than was possible with previous finite-state approaches [45, 46]. The dialogue manager's interpreter finds and returns a dialogue pattern, which is a plan that matches the current context. Recently this has been reimplemented using Prolog in GnuTutor, an open-source approximation of AutoTutor, which allows for more sophisticated backtracking [47]. The dialogue plan returned by the dialogue manager ends with a question, has a gesture, or has rising intonation to indicate to the student that the tutor expects them to respond.

The learning gains of AutoTutor have been evaluated in over 20 experiments conducted during the last 12 years. Assessments of AutoTutor on learning gains have shown effect sizes of approximately 0.8 standard deviation units in the areas of computer literacy [12] and Newtonian physics [13]. AutoTutor's learning gains have varied between 0 and 2.1 sigma (a mean of 0.8), depending on the learning performance measure, the comparison condition, the subject matter, and the version of AutoTutor.

4 Expert human tutoring

Most human tutoring studies that have been reported in peer-reviewed sources have primarily included untrained or "typical" tutors [2]. By comparison, expert tutoring studies are scarce, and such studies have included only a handful of expert tutors. This section reviews the expert tutoring studies most frequently cited in the literature and notes some problems that have contributed to our lack of expert tutoring knowledge.

First, several well-known studies do not mention the number of expert tutors included in the analyses [48, 35, 49, 50]. Second, although some studies report five or six expert tutors [51, 45, 52, 53, 13], many included only one or two experts [21, 54–57]. Third, some of these studies have overlapping expert tutors. For example, the tutors included in [45], [57], and [13] are the same five tutors. Fourth, not all studies on expert human tutoring investigate the same phenomena. A number of studies have focused on the motivational aspects of tutors instead of the cognitive and pedagogical features that contribute to student learning (e.g., the studies by Mark Lepper and colleagues). A fifth problem with these studies is that the credentials of the expert tutors are inconsistent. Some studies define expert tutors as Ph.D.s with extensive teaching or tutoring experience [54, 55, 45, 57], but other studies define expert tutors as graduate students working in tutoring centers [35]. Taken together, these problems raise uncertainty as to whether the findings generalize to all expert tutors. A more detailed analysis of these problems reveals two enduring themes. First, many studies suffer from small or unknown sample sizes. Secondly, it is difficult to aggregate findings across studies because of shared tutors, differing research goals, and inconsistent definitions of expertise.

To address these issues, we recently undertook a rigorous, large scale study of accomplished, expert human tutors. Our approach mirrors our previous study of novice human tutoring by collecting observations of naturalistic one-to-one tutoring. Twelve expert math and science tutors were recruited to participate in the project. The expert tutors were recommended by academic support personnel from public and private schools in a large urban school district. All of the tutors have long-standing relationships with the academic support offices that recommend them to parents and students. The criteria for being an expert tutor in our project are as follows:

- Have a minimum of five years of one-to-one tutoring experience (most of the tutors in our study have 10+ years of tutoring experience)
- Have a secondary teaching license
- Have a degree in the subject that they tutor
- Have an outstanding reputation as a private tutor
- Have an effective track record (i.e., students who work with these tutors show marked improvement in the subject areas for which they receive tutoring)

All of the students in our study were having difficulty in a science or math course and were either recommended for tutoring by school personnel or sought professional tutoring help.

We created our expert tutoring corpus by observing our expert tutors in one-on-one tutoring sessions. Fifty one-hour tutoring sessions were videotaped and transcribed. All of the videotapes were transcribed according to strict transcription guidelines and were verified for accuracy. To capture the complexity of what transpires during a tutoring interaction, three coding schemes were developed to classify every tutor and student dialogue move in the 50 hours of tutoring. In the analyses we conducted, a dialogue move was either a speech act (e.g., a tutor hint), an action (e.g., student reads aloud), or a qualitative contribution made by a student (e.g., partial or vague answer). Multiple dialogue moves could occur within one conversational turn.

Table 3. Tutor Motivational Moves

Move Category	Example
Attribution Acknowledgment	that's easy
Conversational Ok	alrighty
General Motivational Statement	cause you're such a good student I just enjoy ...
Humor	so you're going to have kids and you go "oh I ...
Negative Feedback	no no no no
Negative Feedback Elaborated	actually no you're gonna have some ...
Neutral Feedback	not quite
Neutral Feedback Elaborated	mm you're thinking of vertical vertical angles ...
Positive Feedback	very good alright
Positive Feedback Elaborated	very good because everything is on top
Repetition	negative 2
Solidarity Statement	let's do it

Table 4. Tutor Pedagogical Moves

Move Category	Example
Counter Example	not multiply we'll add in the area of the bases ...
Comprehension Gauging Question	you see what I'm saying
Direct Instruction/Explanation	so that's your lateral area
Example	so as a male you will undergo meiosis and your ...
Forced Choice	so are we going bigger or smaller
Hint	but now we're not gonna add this many dots ...
New Problem	let's look at this example here it's called ...
Other	does he give you a time limit
Paraphrase	you take out an r squared and you'd have 4 ...
Provide Correct Answer	first outer inner last
Preview	we're going to talk about how atoms ions ...
Prompt	can we simplify the radical of 9 is simply
Pump	and then what do we do
Simplified Problem	what inside the cell would have an electrical ...
Summary	so that's all there is to it so you got a circular ...

Two coding schemes were used to classify the tutor dialogue moves, the Tutor Pedagogical Scaffolding scheme and the Tutor Motivational Dialogue scheme. The Pedagogical Scaffolding scheme included 14 categories and was inspired by previous tutoring research on pedagogical strategies and dialogue moves [10, 58]. The Tutor Motivational Dialogue scheme included 13 categories that were either reported previously in the literature or were extrapolated from the INSPIRE model [50]. Each tutor dialogue move was classified as either pedagogical or motivational. The Tutor Motivational and Pedagogical Schemes are presented in Table 3 and Table 4.

A 16 category coding scheme was also developed to classify all student dialogue moves. Some of the student move categories captured the qualitative nature of a student dialogue move (e.g., Correct Answer, Partially Correct Answer, Error-ridden Answer), whereas others were used to classify types of ques-

Table 5. Student Dialogue Moves

Move Category	Example
Acknowledgment	yes, ma'am
Common Ground Question	the parasites?
Correct Answer	6 times 54
Error Ridden Answer	multiply
Gripe	I might as well not pay attention
Knowledge Deficit Question	well, what's a skeleton?
Metacomment	I don't know what I'm doing, hold on
Misconception	I thought you added two to it
No Answer	it will be, oh shoot it will be
Other	she didn't do that
Partial Answer	so I guess eliminate those 2
Read Aloud	first class levers are the most common type a pire of
Social Coordination Action	afternoon sunday? want to do it like sunday afternoon?
Student Works Silently	uh
Think Aloud	to the power of, no, x plus 1
Vague Answer	cause you, yeah times

tions, conversational acknowledgments, and student actions (e.g., reading aloud or solving a problem). The Student Dialogue Move Scheme is presented in Table 5. Four trained judges coded the 50 transcripts on the three dialogue move schemes. Cohen's Kappas were computed to determine the reliability of their judgments. The Kappa scores were .96 for the Tutor Motivational Scheme, .88 for the Tutor Pedagogical Scheme, and .88 for the Student Move Scheme. Approximately 47,000 dialogue moves were coded.

In addition to these dialogue move coding schemes, we also developed a coding scheme for larger units of the tutoring session. We call these units *modes* [14]. Two trained judges coded the 50 transcripts and found eight modes, including *Introduction, Lecture, Highlighting, Modeling, Scaffolding, Fading, Off-Topic,* and *Conclusion*, with Kappa above .80 for each mode. Each mode can be characterized by a specific kind of interaction:

Introduction. Expert tutoring sessions usually begin with an *Introduction* that contains greetings and establishes an agenda for the rest of the session.

Lecture. Approximately 20% of the sessions consist of direct instruction. We call these modes *Lecture*, but they are usually highly customized, just-in-time, and interactive, unlike traditional classroom lecture.

Highlighting. When a student encounters difficulty while problem solving, *Highlighting* draws attention to a problem solving step.

Modeling. In this mode, the tutor works a problem while the student watches.

Scaffolding. Expert human tutoring is dominated by *Scaffolding*, in which the tutor and student solve a problem together. Roughly 50% of all turns take place in this mode.

Fading. As the inverse to *Modeling*, the student predominantly solves a problem while the tutor watches in *Fading*.

Off-Topic. Non-tutoring related conversation, e.g. humor, infrequently occurs.

Conclusion. Expert tutoring sessions usually end in a characteristic fashion, similarly to how *Introduction* begins the session.

An individual mode can span dozens of turns, and so represents a major unit in the structure of a tutoring session. However, not all modes are equally prevalent or contain comparable numbers of turns. Approximately 70% of all turns are contained within *Lecture* and *Scaffolding* modes, with the remaining turns roughly divided amongst the remaining modes [14].

We are currently analyzing these 50 hours of expert human tutoring data, which we call the expert human tutoring corpus (EHTC). Our goal is to explore whether there are similar structures that we have found for novice tutoring, i.e. the 5-Step Tutoring Frame, EMT dialogue, and conversational turn management described in Section 2. However, there may be characteristics of expert human tutoring that are very different than those of novice human tutoring protocols. Although our analyses are still underway, we believe that expert tutors are different from novice tutors in at least six different ways [59–61]. It is important to qualify these claims about expert tutors because with the exception of [61], there was never a systematic comparison of tutors with different expertise in any given study. Instead, the relative frequencies of tutor strategies and discourse moves were computed in the EHTC and compared with the relative frequencies of the same theoretical categories in published studies with unskilled tutors. At the same time, however, there is little evidence of EMT dialogue in the EHTC, and the patterns of dialogue are more complex than the typical instantiation of the 5-Step Tutoring Frame.

1. Expert tutors form more accurate student models that non-expert tutors. This is evidenced in the question asking analyses that we performed. Expert tutors ask proportionately more low specificity questions (e.g., So?) and more common ground questions (e.g., So, I use the Pythagorean Theorem?) than tutors and students in non-expert sessions. We interpret these findings to mean that expert tutors are more attuned to the needs of their students and have established considerable common ground. If this wasn't the case, low specificity questions (e.g., So?) would result in conversation breakdowns. We also found that students being tutored by experts ask fewer knowledge deficit questions (e.g., What do the ribosomes do?) than students working with non-expert tutors, indicating that knowledge deficit questions are less necessary when participants have established a high level of common ground.
2. Expert tutors are more dynamic in their instruction and do not rely on curriculum scripts. Experts typically begin the tutoring sessions by figuring out the topics/problems that students are having difficulty with and by asking questions about the students' performance on quizzes, homework, and exams. After this data collection phase, the tutor decides where to begin the session and what material will be covered. Expert tutors do not begin a session with any pre-planned teaching agenda, but rather base their instruction on students' particular needs at the time of the tutoring session.

3. Expert tutors give more discriminating feedback than non-expert tutors. Non-experts are just as likely to give positive feedback to wrong answers as negative feedback [11], but this is not true of expert tutors.
4. Expert tutors primarily rely on just-in-time direct instruction and evaluative feedback when responding to student dialog moves.
5. Expert tutors are task-oriented, direct, and do not appear to adhere to Lepper INSPIRE motivational model.
6. Particular tutoring modes (defined by tutor dialogue move frequencies and patterns) are evident in expert tutoring, including *Introduction, Lecture, Highlighting, Modeling, Scaffolding, Fading, Off-Topic,* and *Conclusion.*

We have recently used data mining techniques to discover significant patterns of dialogue moves in the EHTC [15]. Our basic approach is to consider two-step transitions, i.e. move to move, that significantly diverged from chance and whose effect sizes were greater than the median effect size. So far our analyses have focused on *Lecture*. In *Lecture*, only 34 transitions out of 1869 (43 x 43) are significant with effect sizes above the median. A visual inspection of these transitions revealed four meaningful clusters. In the first cluster, the information *transmission* cluster, the tutor mostly engages in direct instruction and only superficially monitors student attention and understanding. In the second cluster, the information *elicitation* cluster, the tutor elicits information from the student using direct questioning, e.g. forced choice, prompts, pumps, etc., the student tries to answer, and the tutor gives feedback on the student's answer. The information elicitation cluster is the *Lecture* cluster most like the IRF and 5-Step Tutoring Frame described in Section 2. The third cluster is the off-topic cluster, e.g. humor, consisting of just a few moves as opposed to the *Off-Topic* mode. The fourth and final cluster in *Lecture* is the questioning cluster that handles student-asked common ground questions and knowledge deficit questions. Each of these four clusters can be viewed as a subgraph of the larger *Lecture* graph or viewed as a subdialogue nested in the larger *Lecture* dialogue.

Our analyses of the EHTC have revealed a richer structure than has previously been reported for novice tutoring, though again, we stress that this might be confounded by the lack of novice tutors in our sample. The behavior of novice tutors, as described in Section 2, aligns fairly well with the distinction of inner and outer loop in the behavior of ITS [62]. According to the inner/outer loop distinction, problem selection happens in the outer loop, and the actual working of the problem, step by step, happens in the inner loop. Under this analysis, the three features of novice tutoring described in Section 2 align quite well with the inner loop. The 5-Step Tutoring Frame provides an overall dialogue structure for the inner loop, with step 4 accounting for much of the individual steps, or expectations, in the problem. EMT dialogue contributes by structuring the content within step 4. Finally, the conversational turn management further elaborates step 4. As stated in Section 2, these three structures are multiply embedded.

However, the multiple levels of structure found in our analysis of the EHTC, while embedded, are considerably more complex. Transition probabilities between modes indicate that sessions typically shift from *Introduction* to *Lecture* to *Scaffolding*. Because only some modes introduce problems, e.g. *Scaffolding*, mode transitions are a step above the outer loop of problem selection. Inner loops occur in clusters within modes, such as the information elicitation cluster in *Lecture*. Finally, there are the dialogue moves themselves. In contrast to the more straightforward multiple embedding of the 5-Step Tutoring Frame, EMT dialogue, and conversational turn management, the EHTC corpus is revealing a complex embedding in which many clusters exist in a single mode, and many dialogue moves exist within each cluster. In other words, the EHTC is revealing a web of embedded structure in expert human tutoring, as opposed to the simple nesting found in novice human tutoring.

5 Guru

Guru, like AutoTutor, is designed to simulate a human tutor by holding a conversation with the learner in natural language. However, Guru is design to simulate an *expert* human tutor rather than a novice human tutor. The characteristics of expert human tutors described in Section 4 are informative when considering the design of Guru, and how it should differ from AutoTutor.

First and foremost, our analyses revealed that expert tutors do not use curriculum scripts. However, curriculum scripts, are a central element of AutoTutor. They contain all the EMT dialogue for a problem as well as the expectations which are used to track the student's progress and understanding. If curriculum scripts and EMT dialogue are not characteristic of expert human tutoring, then Guru requires a new way of tracking student understanding and organizing knowledge about the domain.

Second, in terms of dialogue structure, expert tutors rely a great deal on evaluative feedback and just-in-time direct instruction. Contrast this to the hints, prompts, and elaborations that constitute the bulk of AutoTutor's dialogue. Guru cannot soley rely on hints, prompts, and elaborations but rather must incorporate tutor dialogue moves into a new model for just-in-time direct instruction.

Third, experts are precise with their feedback. In AutoTutor, feedback is calculated by comparing the student's responses with the expectations from the curriculum script. Again, expert tutors do not appear to use such a script. Furthermore, the traditional way of comparing student answers with expectations in AutoTutor, LSA [12, 13, 43], is relatively imprecise: to LSA, "do you want to drive me" and "do you want me to drive" mean the same thing. To model the precise feedback of expert human tutors, it is necessary to incorporate a more sensitive technique than LSA.

Fourth, expert tutors maintain highly accurate student models. In AutoTutor, the student model is simply the set of LSA comparisons of the student's input to each expectation in the curriculum script. Not only do expert tutors not use curriculum scripts, but LSA also doesn't have the precision to match an expert tutor. Therefore Guru should apply a different methodology for student modeling.

Fifth, expert tutors use a variety of tutoring modes and clusters within modes that have no clear correlates in AutoTutor. Contrasted with the linear hint-prompt-assertion cycle used in AutoTutor, the expert tutoring modes are both more numerous and more complex. Fortunately, the dialogue management used in AutoTutor is extremely powerful [18], so a new approach to dialogue management per se for Guru is not required.

In summary, Guru needs a new way to model the domain, model the student, interpret student utterances, and generate direct instruction. We are working on a unified approach to all of these tasks, which is based on a single knowledge representation. Using a single knowledge representation for multiple purposes like these is not uncommon in an ITS. For example, overlay student models typically assume a domain decomposition in which chunks of content can be marked as understood by the student, rather like checking items off a list. An overlay student model is so called because it lays over the domain model in a rather transparent way, i.e. each element of the domain model is on the checklist for the overlay student model.

Clearly an overlay student model first requires a domain model. In the same way, interpretation of student input and the generation of direct instruction can also be yoked to a domain model. However, the creation of a domain model is sufficiently challenging to require special authoring tools and many man-hours to develop [63–65]. Thus for Guru we have been particularly interested in unsupervised and semi-supervised knowledge representation techniques that can extract semantic representations from raw text. Although we still find LSA useful for some tasks, we have been developing a new technique for concept map extraction, which we believe holds promise for domain modeling, student modeling, interpretation of student utterances, and generation of direct instruction.

5.1 Concept map extraction

The term "concept map" has become largely associated with an educational practice in which students create a graph representation of ideas and the links between them [66]. However, similar notions to concept maps have been used in the education, artificial intelligence, and psychological communities for decades, and as a result there are dozens of different definitions of concept map [67]. Generally speaking, a concept map consists of a set of nodes (concepts) and edges (relations) describing a core concept or answering a core question [66]. We call a pair of nodes connected by an edge a *triple* because it consists of three elements: a start node, an edge relation, and an end node. Thus in general, relationships in concept maps are binary. This prevents or obfuscates the expression of some relationships such as a verb with three arguments, unless additional constraints are adopted which can convert a concept map into a first order logic [68, 69].

Our unique concept map definition is a synthesis of previous work in both the psychology and education literatures [70, 71, 67]. The education literature, particularly relevant from an ITS point of view, has promoted relatively small, human-readable maps, such as the SemNet map [67]. The key feature that makes these concept maps easy to understand is that they are radial, with a core concept in the middle of the map and a single layer of links radiating from that concept. End nodes linked to the core concept can potentially be the centers of their own maps, but each map is coherent by itself. From the psychology literature, we adopt a limited set of edges linking two nodes in the concept map [70, 71]. Discrete sets of edges are also common in ontologies, e.g. `is-a` or `has-part`. For Guru, a salient advantage of having a restricted set of edges is that they facilitate both generating questions and answering questions from the map [70, 71].

Recently, we developed a semi-supervised procedure for extracting concept maps with radial structure and discrete set of edges [72]. The procedure operates on a textbook, using a semantic parser and post processing to transform the semantic parse into concept maps. More specifically, the LTH SRL Parser [73] outputs a dependency parse annotated with semantic roles derived from Propbank [74] and Nombank [75] for each sentence in the textbook. For each syntactic or semantic relation found by the parser, we require that the start node be a key term in our domain. Key terms are defined as those terms existing in the glossary or index of the book. If the start node is a key term, a corresponding end term is found in the parse, and then the relation linking them is classified using a hand-built decision tree. Some relations are syntactic, e.g. `is-a` is determined by the presence of a "be" main verb as in "an abdomen is the posterior part of an arthropod's body." Other relations are semantic and are classified using the semantic information returned by Nombank or Propbank, e.g. `has-part` is determined by "body" in the example above because "body" is a Nombank predicate whose sense gloss is "partitive part." This process of concept map extraction is semi-supervised because the key terms and edge relations have been manually defined for our domain, but the rest of the procedure is unsupervised.

5.2 Domain and student modeling

The concept map extracted from Guru's biology textbook contains roughly 30,000 triples centered around 2,000 terms. Thus it is a fairly well elaborated model of the domain. The triples allow us to query particular properties of the key terms in our domain:

ABDOMEN is-a part
ARTHROPOD has-part ABDOMEN
ABDOMEN has-property posterior

It is fairly straightforward to build an overlay student model around this domain model. One can consider each key term as a chunk the student should master, and calculate a coverage score based on the number of triples a student has appeared to master. Although each chunk may be considered as a kind of expectation, or bundle of expectations, the overall structure of the concept map-based domain model is different from the script based model of EMT dialogue described in Section 2, in at least three ways. First, the concept map expectations are not attached to a particular problem, but instead are general to the domain. Second, rather than a limited set of expectations, the concept map (in theory) includes all of the salient relations in the biology textbook. Finally, the concept map relations, consisting of triples, are more structured than AutoTutor expectations, which are undifferentiated LSA vectors. In other words the concept map-based domain model appears to be more general, have broader coverage, and be more structured than curriculum script based EMT dialogue.

We are currently building richer links between the standards for high school biology instruction in our state and concept maps we've extracted from the state textbook. This will allow us to better focus the domain and student models of Guru to the content covered by state-wide standardized testing, which in turn will make it easier to integrate Guru into classroom activities.

5.3 Interpretation of student utterances

In Guru, and in an ITS generally, interpreting a student utterance means mapping that utterance to the domain model. In the case of Guru, which uses an overlay student model, such mapping facilitates both interpretation of the student utterance as well as assessment of the student's current understanding. The most straightforward way to accomplish this mapping is to use the concept map extraction technique from Section 5.1 on the student's utterances, and compare the resulting triples with those in the domain and student models. Intuitively, there are more ways to compare triples than monolithic LSA vectors. By definition, each triple has three components, and Guru's feedback can be differentially driven by the correctness of each component.

If only the start node of a student's triple is incorrect, we can hypothesize that the student has not adequately discriminated their start node from the actual start node. For example, if the student's utterance contains the triple `white blood cell` *has-consequence* `delivers oxygen`, then we can identify that this student knows something about red blood cells that is being incorrectly generalized to white blood cells. If only the edge relation of a student's triple is incorrect, then we can hypothesize that the student knows two concepts are related, but misunderstands the type of relation. For example, if the student's utterance contains the triple `red blood cell` *lacks* `delivers oxygen`, then we can target the *lacks* relation for remediation. Finally, if the end node only is incorrect, then we might hypothesize that the student lacks sufficient background knowledge. For example, the triple `red blood cell` *has-property* `found in plants` likely indicates that the student knows absolutely nothing about red blood cells, and some direct instruction is needed. These are just examples of possible strategies, but they illustrate how a concept map representation composed of triples can be used to make fine discriminations of the student's error and respond appropriately.

5.4 Generating direct instruction

Just-in-time direct instruction is, by definition, unplanned. As such it is impossible to render from a curriculum script, which is essentially pre-planned. Rather just-in-time direct instruction must be generated dynamically from an existing domain model. Concept maps have been previously used to generate text. Our concept maps use a fixed set of edge relations that can be set into correspondence with certain question types, e.g. definitional, causal consequent, and procedural, for both the purposes of answering questions [70] as well as generating them [71]. For example, `red blood cell` *has-consequence* `delivers oxygen` can be used to generate the questions "What causes oxygen to be delivered," "What does a red blood cell do," or "What can you say about a red blood cell and oxygen" depending on whether we want to query the start node, the end node, or the edge relation between them respectively. Of course, given the same triple, it is straightforward to create direct instruction like "a red blood cell delivers oxygen."

A similar approach is used in the Betty's Brain ITS [76, 77]. In this "learning by teaching" system, students teach an animated agent named Betty, whose brain is visible as a causal concept map with additional hierarchical (i.e. is-a) and descriptive relations (i.e. has-property). Students teach Betty by explicitly creating linkages in the concept map "brain." Betty can "take" quizzes by applying a qualitative reasoning algorithm to the causal concept map. Moreover, Betty can describe her reasoning by reading off the relationships in the map, e.g. light increases algae growth which decreases oxygen in the water.

5.5 Limitations

In this section, we have outlined some of the major dimensions in which we believe expert human tutors differ from novice human tutors, and the implications for these differences on the design of an expert ITS. The major differences that we have emphasized, the domain model, the student model, interpretation of student utterances, and generation of direct instruction, appear to be well supported by a concept map knowledge representations. However, although our preliminary observations are plausible, these applications have yet to be rigorously evaluated.

6 Conclusion

This chapter described our previous and ongoing investigations into the conversational interaction that defines human tutoring. Both our analyses of novice and expert human tutors are corpus-based, driven by extensive collections of human tutoring dialogues. Our goal is to better understand the representations and processes of human tutoring by building computational models in the form of intelligent tutoring systems that embody our theory.

For novice human tutoring, we have identified three major dialogue structures, including the 5-Step Tutoring Frame, EMT dialogue, and conversational turn management. These three structures are nested such that each occurs within its preceding structure. These three structures are comprehensive enough that they can be used to specify the runtime of an ITS, and we have done so in the ITS

AutoTutor. The 5-Step Tutoring Frame defines the overall structure of a problem, the EMT dialogue defines the components of a problem, and conversational turn management defines how each tutor turn is constructed in a conversationally appropriate way. In experimental evaluations of learning gains, AutoTutor yields an approximately .8 effect size increase relative to control conditions. Relative to the .4 effect size for novice human tutoring reported in a meta-analysis [2], AutoTutor appears to be convincing as a model of novice human tutoring both in terms of its structure and its effectiveness.

Our recent collection and analysis of expert human tutoring has revealed some differences which may be attributable to the difference between expert and novice human tutors. The expert tutors in our study manifested very complex conversational interaction relative to novice human tutors, including dialogue modes, functional clusters of dialogue moves within modes, and finally the dialogue moves themselves. As discussed in Section 4, there is no clear correspondence between these dialogue structures and the structures associated with novice human tutoring. Moreover, our analyses of expert human tutoring suggest that expert human tutors utilize more precisely defined and well-organized domain and student models, are more precise in evaluating and responding to student answers, and utilize a just-in-time direct instruction that is highly adapted to the student's current knowledge state.

We have proposed a particular formulation of concept maps to address these four issues, and we have outlined how these concept maps can be extracted from a textbook, alleviating the burden of domain model authoring. Using the concept map representation, we have further proposed several strategies for addressing the four salient phenomena in our analysis of expert human tutoring, including the domain/student model, interpretation of student utterances, and generation of direct instruction. Though our current assessments are promising, these strategies await a more rigorous evaluation.

Moreover, since the goal of any ITS is to produce learning gains, the conclusive evaluation of the concept map representation and associated strategies is a learning outcome study. We are currently engaged in curriculum development, usability studies, and unit testing in preparation for a learning outcome study. If we have properly identified and represented the differences between expert and novice human tutors, then this should be reflected in a corresponding difference in learning gains.

References

1. Bloom, B.S.: The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher **13**(6) (June 1984) 4–16
2. Cohen, P.A., Kulik, J.A., Kulik, C.L.C.: Educational outcomes of tutoring: a meta analysis of findings. American Educational Research Journal **19** (1982) 237–248
3. Corbett, A.: Cognitive computer tutors: Solving the two-sigma problem. In: User Modeling: Proceedings of the Eighth International Conference, Springer (2001) 137–147

4. Graesser, A.C., Person, N.K.: Question asking during tutoring. American Educational Research Journal **31**(1) (1994) 104–137
5. Chi, M.T.H., Siler, S.A., Jeong, H., Yamauchi, T., Hausmann, R.G.: Learning from tutoring. Cognitive Science **25** (2001) 471–533
6. Chi, M., Roy, M., Hausmann, R.: Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. Cognitive Science **32**(2) (2008) 301–341
7. Wiley, J., Jensen, M.: When three heads are better than two. In: Proceedings of the 28th annual conference of the cognitive science society. (2006) 2375–2380
8. Wiley, J., Bailey, J.: Effects of collaboration and argumentation on learning from web pages. In M., O.A., C.E., H., Erkens, G., eds.: Collaborative learning, reasoning, and technology. Lawrence Earlbaum, Mahwah, New Jersey (2006) 297321
9. Palinscar, A.S., Brown, A.: Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction **1** (1984) 117–175
10. Graesser, A.C., Person, N.K., Magliano, J.P.: Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied Cognitive Psychology **9** (1995) 1–28
11. Person, N.K., Graesser, A.C., Magliano, J.P., Kreuz, R.J.: Inferring what the student knows in one-to-one tutoring: the role of student questions and answers. Learning and individual differences **6**(2) (1994) 205–229
12. Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., Louwerse, M.M.: AutoTutor: A tutor with dialogue in natural language. Behavioral Research Methods, Instruments, and Computers **36** (2004) 180–193
13. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rose, C.: When are tutorial dialogues more effective than reading? Cognitive Science **31** (2007) 3–62
14. Cade, W.L., Copeland, J.L., Person, N.K., D'Mello, S.K.: Dialogue modes in expert tutoring. In: ITS '08: Proceedings of the 9th international conference on Intelligent Tutoring Systems, Berlin, Heidelberg, Springer-Verlag (2008) 470–479
15. D'Mello, S., Olney, A.M., Person, N.: Mining collaborative patterns in tutorial dialogues. Journal of Educational Data Mining (in press)
16. Person, N., Graesser, A.: Evolution of discourse during cross-age tutoring. In O'Donnell, A.M., King, A., eds.: Cognitive perspectives on peer learning. Lawrence Erlbaum, Mahweh, New Jersey (1999) 69–86
17. Person, N.K., Kreuz, R.J., Zwaan, R., Graesser, A.C.: Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. Cognition and Instruction **13** (1995) 161–188
18. Graesser, A.C., Chipman, P., Haynes, B., Olney, A.: AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education **48**(4) (November 2005) 612– 618
19. Chi, M.T.H., Siler, S.A., Jeong, H.: Can tutors monitor students' understanding accurately? Cognition and Instruction **22**(3) (2004) 363–387
20. Litman, D., Forbes-riley, K.: Correlations between dialogue acts and learning in spoken tutoring dialogues. Nat. Lang. Eng. **12**(2) (2006) 161–176
21. Shah, F., Evens, M., Michael, J., Rovick, A.: Classifying Student Initiatives and Tutor Responses in Human Keyboard-to-Keyboard Tutoring Sessions. Discourse Processes **33**(1) (2002) 23–52
22. Sinclair, J., Coulthard, R.M.: Towards an Analysis of Discourse: the English used by teachers and pupils. Oxford University Press, London (1975)
23. Graesser, A.C., DMello, S.K., Cade, W.L.: Instruction based on tutoring. In Mayer, R., Alexander, P., eds.: Handbook of Research on Learning and Instruction. Routledge, New York (in press)

24. Person, N., Graesser, A.: Fourteen facts about human tutoring: Food for thought for ITS developers. In: AI-ED 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a View Toward the Classroom. (2003) 335–344
25. Collins, A., Warnock, E.H., Aiello, N., Miller, M.L.: Reasoning from incomplete knowledge. In Bobrow, D., Collins, A., eds.: Representation and Understanding: Studies in Cognitive Science. Academic Press, New York (1975)
26. Rogoff, B., Gardner, W.: Adult guidance of cognitive development. In Rogoff, B., Lave, J., eds.: Everyday cognition: Its development in social context. Harvard University Press, Cambridge, MA (1984) 95–116
27. Sleeman, D., Brown, J.S., eds.: Intelligent tutoring systems. Academic Press, New York (1982)
28. Gagn, R.M.: The conditions of learning and theory of instruction. Holt, Rinehart and Winston (1985)
29. Lesgold, A., Lajoie, S., Bunzo, M., Eggan, G.: SHERLOCK: A coached practice environment for an electronics troubleshooting job. In Larkin, J.H., Chabay, R.W., eds.: Computer assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches. Lawrence Earlbaum, Hillsdale, NJ (1992) 201–238
30. Aleven, V., Koedinger, K.R.: An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. Cognitive Science **26** (2002) 147–179
31. Chi, M., Leeuw, N., Chiu, M., LaVancher, C.: Eliciting Self-Explanations Improves Understanding. Cognitive Science **18**(3) (1994) 439–477
32. VanLehn, K., Jones, R., Chi, M.: A Model of the Self-Explanation Effect. The Journal of the Learning Sciences **2**(1) (1992) 1–59
33. Anderson, J.R., Corbett, A.T., Koedinger, K.R., Pelletier, R.: Cognitive tutors: lessons learned. The Journal of the Learning Sciences **4**(2) (1995) 167–207
34. VanLehn, K., Jordan, P., Rose, C., Bhembe, D., Bottner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberg, M., Roque, A., et al.: The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In: Proceedings of Intelligent Tutoring Systems Conference. Volume 2363., Springer (2002) 158–167
35. Fox, B.: The human tutoring dialogue project. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ (1993)
36. Moore, J.D.: Participating in explanatory dialogues: Interpreting and responding to questions in context. The MIT Press, Cambridge, MA (1994)
37. Jackson, G.T., Person, N.K., Graesser, A.C.: Adaptive tutorial dialogue in autotutor. In: Proceedings of the workshop on Dialog-based Intelligent Tutoring Systems at the 7th International conference on Intelligent Tutoring Systems, Universidade Federal de Alagoas, Brazil (2004) 368–372
38. Deerwester, S.C., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.A.: Indexing by latent semantic analysis. Journal of the American Society of Information Science **41**(6) (1990) 391–407
39. Dumais, S.: Improving the retrieval of information from external sources. Behavior Research Methods, Instruments and Computers **23**(2) (1991) 229–236
40. Landauer, T.K., Dumais, S.T.: A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review **104** (1997) 211–240
41. Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W., eds.: Handbook of Latent Semantic Analysis. Lawrence Erlbaum, Mahwah, New Jersey (2007)
42. Olney, A., Louwerse, M., Mathews, E., Marineau, J., Hite-Mitchell, H., Graesser, A.: Utterance classification in AutoTutor. In: Proceedings of the HLT-NAACL

03 Workshop on Building Educational Applications Using Natural Language Processing, Philadelphia, Association for Computational Linguistics (2003) 1–8
43. Graesser, A.C., Olney, A.M., Haynes, B.C., Chipman, P.: AutoTutor: A cognitive system that simulates a tutor that facilitates learning through mixed-initiaive dialogue. In Forsythe, C., Bernard, M.L., Goldsmith, T.E., eds.: Cognitive Systems: Human Cognitive Models in Systems Design. Erlbaum, Mahwah, NJ (2005) 177–212
44. Larsson, S., Traum, D.: Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. Natural Language Engineering (2000) 323–340 Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering.
45. Graesser, A.C., Person, N.K., Harter, D., et al.: Teaching tactics and dialog in AutoTutor. International Journal of Artificial Intelligence in Education **12**(3) (2001) 257–279
46. McTear, M.: Modelling spoken dialogues with state transition diagrams: experiences with the CSLU toolkit. In: Fifth International Conference on Spoken Language Processing, Sydney, Australia, ISCA (December 1998)
47. Olney, A.M.: Gnututor: An open source intelligent tutoring. system based on AutoTutor. In: Proceedings of the 2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems, Washington, DC, AAAI Press (November 5-7 2009) 70–75 FS-09-02.
48. Aronson, J.: Improving academic achievement: Impact of physchological factors on education. Academic Press, San Diego, CA (2002)
49. Derry, S.J., Lajoie, S.P.: Computers as cognitive tools. Erlbaum, Hillsdale, NJ (1993)
50. Lepper, M.R., Woolverton, M.: The wisdom of practice: Lessons learned from the study of highly effective tutors. Improving academic achievement: Impact of psychological factors on education. Academic Press, San Diego, CA (2002) 135–158
51. Derry, S.J., Potts, M.K.: How tutors model students: A study of personal constructs in adaptive tutoring. American Educational Research Journal **35**(1) (1998) 65–99
52. Lepper, M.R., Aspinwall, L.G., Mumme, D.L., Chabay, R.W.: Self-perception and social-perception processes in tutoring: Subtle social control strategies of expert tutors. In Olson, J.M., Zanna, M.P., eds.: Self-inference processes: The Ontario symposium. Erlbaum, Hillsdale, NJ (1990) 217–237
53. Lepper, M.R., Woolverton, M., Mumme, D., Gurtner, J.: Motivational techniques of expert human tutors: Lessons for the design of computer based tutors. In Lajoie, S., Derry, S., eds.: Computers as cognitive tools. Erlbaum, Hillsdale, NJ (1993)
54. Evens, M., Spitkovsky, J., Boyle, P., Michael, J., Rovick, A.: Synthesizing tutorial dialogues. In: Proceedings of the 15th Annual Conference of the Cognitive Science Society, Lawrence Erlbaum Associates (1993)
55. Glass, M., Kim, J., Evens, M., Michael, J., Rovick, A.: Novice vs. expert tutors: A comparison of style. In: Midwest Artificial Intelligence and Cognitive Science Conference. (1999)
56. Lajoie, S., Faremo, S., Wiseman, J.: Tutoring strategies for effective instruction in internal medicine. International Journal of Artificial Intelligence and Education **12** (2001) 293–309
57. Jordan, P., Siler, S.: Student initiative and questioning strategies in computer-mediated human tutoring dialogues. In: Proceedings of ITS 2002 Workshop on Empirical Methods for Tutorial Dialogue Systems. (2002)
58. Cromley, J.G., Azevedo, R.: What do reading tutors do? a naturalistic study of more and less experienced tutors in reading. Discourse Processes **40**(2) (2005) 83–113

59. Person, N.K., Lehman, B., Ozbun, R.: Pedagogical and motivational dialogue moves used by expert tutors. Presented at the 17th Annual Meeting of the Society for Text and Discourse.Glasgow, Scotland. (July 2007)
60. Person, N.K., Mallott, L.: Question generation in expert tutoring. Presented at the 17th Annual Meeting of the Society for Text and Discourse. Glasgow, Scotland. (July 2007)
61. Person, N.K., Moses, J., Willson, W.: Question asking in expert and non-expert tutoring sessions. Presented at the 2007 AERA Annual Meeting. Chicago, IL. (April 2007)
62. Vanlehn, K.: The behavior of tutoring systems. Int. J. Artif. Intell. Ed. **16**(3) (2006) 227–265
63. Murray, T.: Authoring Knowledge-Based tutors: Tools for content, instructional strategy, student model, and interface design. Journal of the Learning Sciences **7**(1) (1998) 5
64. Corbett, A.: Cognitive tutor algebra I: Adaptive student modeling in widespread classroom use. In: Technology and Assessment: Thinking Ahead. Proceedings from a Workshop, Washington, D.C., National Academy Press (November 2002) 50–62
65. Aleven, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: The cognitive tutor authoring tools (CTAT): Preliminary evaluation of efficiency gains. In: Intelligent Tutoring Systems. (2006) 61–70
66. Novak, J.D., Canas, A.J.: The theory underlying concept maps and how to construct them. Technical report, Institute for Human and Machine Cognition (January 2006)
67. Fisher, K., Wandersee, J., Moody, D.: Mapping biology knowledge. Kluwer Academic Pub (2000)
68. Sowa, J.F.: Conceptual graphs. In Van Harmelen, F., Lifschitz, V., Porter, B., eds.: Handbook of knowledge representation. Elsevier Science, San Diego, USA (2007) 213–237
69. Sowa, J.F.: Concept mapping (March 2010)
70. Graesser, A.C., Franklin, S.P.: Quest: A cognitive model of question answering. Discourse Processes **13** (1990) 279–303
71. Gordon, S., Schmierer, K., Gill, R.: Conceptual graph analysis: Knowledge acquisition for instructional system design. Human Factors: The Journal of the Human Factors and Ergonomics Society **35**(3) (1993) 459–481
72. Olney, A.M.: Extraction of concept maps from textbooks for domain modeling. In: Proceedings of the International Conference on Intelligent Tutoring Systems. (in press)
73. Johansson, R., Nugues, P.: Dependency-based syntactic-semantic analysis with PropBank and NomBank. In: CoNLL '08: Proceedings of the Twelfth Conference on Computational Natural Language Learning, Morristown, NJ, USA, Association for Computational Linguistics (2008) 183–187
74. Palmer, M., Gildea, D., Kingsbury, P.: The Proposition Bank: An annotated corpus of semantic roles. Comput. Linguist. **31**(1) (2005) 71–106
75. Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., Grishman, R.: The NomBank project: An interim report. In Meyers, A., ed.: HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation, Boston, Massachusetts, USA, Association for Computational Linguistics (May 2 - May 7 2004) 24–31
76. Biswas, G., Schwartz, D., Leelawong, K., Vye, N.: Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence **19** (March 2005) 363–392

77. Leelawong, K., Biswas, G.: Designing learning by teaching agents: The Betty's Brain system. Int. J. Artif. Intell. Ed. **18**(3) (2008) 181–208