

Classifying Student Initiatives and Tutor Responses in Human Keyboard-to-Keyboard Tutoring Sessions

Farhana Shah

Department of Computer Science
Quaid-I-Azam University

Martha Evens

Department of Computer Science
Illinois Institute of Technology

Joel Michael and Allen Rovick

Department of Molecular Biophysics and Physiology
Rush Medical College

Please send all correspondence to:

Professor Martha Evens, CS, IIT,

Room 236, 10 West 31st Street

Chicago, IL 60616,

tel: 312-567-5153, fax: 312-567-5067, email: evens@iit.edu

Abstract

This study analyzed 28 hour-long tutoring sessions that were carried out keyboard-to-keyboard with tutor and student in different rooms. The tutors were professors of physiology at Rush Medical College. The students were first year medical students.

We classified student initiatives and tutor responses in human tutoring sessions with the goal of making our intelligent tutoring system capable of handling mixed-initiative dialogue. Student initiatives were classified along four dimensions: communicative goal, surface form, focus of attention, and degree of certainty (does the student hedge or not?). Student goals include: request for confirmation, request for information, challenge, refusal to answer, and conversational repair. Tutor responses were classified along three dimensions: communicative goal, surface form, and delivery mode. The tutor goals included: causal explanation, acknowledgment, conversational repair, instruction in the rules of the game, teaching the problem-solving algorithm, and teaching the language of physiology. Our inter-rater reliability studies supported these categories in the domain of tutoring.

Classifying Student Initiatives and Tutor Responses in Human Keyboard-to-Keyboard Tutoring Sessions

This paper analyzes student initiatives and tutor responses in 28 hour-long human tutoring sessions. The tutors were our collaborators, Joel Michael and Allen Rovick, Professors of Physiology at Rush Medical College. The students were medical students taking the first-year physiology course. This study was motivated both by our interest in tutoring dialogues and our desire to make our intelligent tutoring system, CIRCSIM-Tutor, able to understand student initiatives and respond to them (Woo, Evens, Michael, & Rovick, 1991; Michael, Rovick, Evens, Shim, Woo, & Kim, 1992).

We argue that it is particularly important to study tutoring dialogues. Tutoring is the most effective kind of education known (Bloom et al., 1984). Wells (1999) has shown that tutoring dialogues and language-learning dialogues between parents and children have much in common. But these educational dialogues are very different from other types of dialogue (Person, Kreuz, Zwaan & Graesser, 1995). Most noticeably, these dialogues do not follow the conversational maxims of the philosopher H. Paul Grice (1975). In educational dialogues, it is perfectly appropriate for the teacher to ask questions while knowing the answer. In educational dialogues hinting is a helpful strategy not a hurtful one (Hume, Michael, Rovick & Evens, 1996),

The CIRCSIM-Tutor project began as a cooperative effort between the Computer Science Department at Illinois Institute of Technology and the Department of Physiology at Rush Medical College to build an intelligent tutoring system that carries on a natural language

dialogue capable of teaching students to use causal reasoning to solve problems in physiology. From the beginning we decided to base the design of the system on human tutoring sessions carried out by Joel Michael and Allen Rovick, who were already tutoring students regularly on a face-to-face basis. They switched to keyboard-to-keyboard tutoring so that the system builders would have models of keyboard language. We built CDS (the Computer Dialogue System) to support the collection of these dialogues (Li, Seu, Evens, Michael & Rovick, 1992). CDS allows two PC users to communicate with each other via modems and a telephone line. The input is communicated through the keyboard; it appears on both computer screens. A transcript of the dialogue is saved as a file on the hard disk of one of the computers. We implemented an explicit turn-taking mechanism to avoid the confusion created when both participants type at once. This mechanism requires one side to relinquish the turn before the other starts to type; both sides can request an interrupt but the other participant has to agree before the interrupt takes place. We have now rewritten this program in C under Windows to run on any PC. Anyone who wants a copy should notify the second author.

The sessions all begin with the tutor describing a patient with a pacemaker malfunction, which perturbs the baroreceptor reflex system responsible for preserving a constant blood pressure in the human body. The tutor then asks the student to make predictions about qualitative changes in certain important physiological parameters, first in the Direct Response (DR) period, then in the Reflex Response (RR) period, and finally in Steady State (SS). When the student makes an error, the tutor carries out a remedial dialogue. As one might expect in any technical sublanguage, both the tutor and the student use abbreviations very often, especially for the names of the parameters. Although these names are not important to understanding the points that we

are trying to make about tutoring dialogues in the present study, we list them here for the benefit of any readers who may need clarification.

CC	Cardiac Contractility
CO	Cardiac Output
HR	Heart Rate
MAP	Mean Arterial Pressure
SV	Stroke Volume
RAP	Right Atrial Pressure
TPR	Total Peripheral Resistance

We have already demonstrated that the tutoring done by Michael and Rovick is effective in the sense that tutored students solved problems in causal reasoning on a post test significantly better than classmates who read appropriate text for the same period of time (Michael & Rovick, 1993). We have also collected 31 keyboard-to-keyboard sessions with novice tutors, second year medical students tutoring first year medical students. We found that students tutored by Michael and Rovick achieved significantly better learning outcomes on post-tests than those tutored by novice tutors (Glass, Kim, Evens, Michael & Rovick, 1999).

From the beginning of our work together, Michael and Rovick have stressed the importance of student initiatives. One major argument for building a natural language-based tutor was the desire to understand student explanations whether elicited by the tutor or volunteered by the student. Student initiatives are often volunteered explanations. We believe that these initiatives are the conversational reflection of the self-explanations shown by Chi,

Leeuw, Chiu, and Lavancher (1994) to be a crucial part of learning. Other initiatives are questions from the student. Graesser and Person (1994) have demonstrated the importance of questions in learning new material, although student questions are rare and need to be encouraged in learning environments.

We believe that it is crucially important for our tutor to recognize student initiatives and generate appropriate responses. As we will elaborate below, we have classified student initiatives along four dimensions: the communicative goal, the content area, the surface form, and the degree of certainty expressed (does the student hedge or not). The tutor responses are classified in three dimensions: the communicative goal, the delivery mode, and the surface form. We have defined five communicative goals in the student initiatives and nine in the tutor responses. The student goals are request for confirmation, request for information, challenge, refusal to answer, and conversational repair. The tutor goals are explanation, acknowledgement, conversational repair, instruction in the rules of the game, teaching the problem solving process, help in response to pause, brushing off (changing the subject back to the tutor's agenda), probing the student's inference process, and teaching the sublanguage. The present study reports inter-rater reliability analyses that confirm the validity of these categorization schemes.

We define a *student initiative* as any student contribution to the dialogue that is not an answer to a question asked by the tutor. A *response* is the tutor's reaction to an initiative. In the sessions that we have studied the tutor always responds in some way before going back to his own agenda. Walker and Whittaker (1990, p. 70) relate initiative to control: "As INITIATIVE passes back and forth between the discourse participants, we say that CONTROL over the conversation gets transferred from one discourse participant to another." They analyze two different types of dialogues: task-oriented dialogues and advice-giving dialogues.

Advice-giving dialogues seem closer to tutorial dialogues but there are major differences. Turns tend to be much shorter in our tutorial dialogues than in those studied by Walker and Whittaker. Another important difference between our tutorial dialogues and the dialogues studied by Walker and Whittaker is that our dialogues were collected using a keyboard-to-keyboard system, which has an explicit mechanism for relinquishing the turn. We are perhaps even closer to Green and Carberry (1999), who view the initiative in dialogue in terms of a speaker taking the chance of saying more than was her/his obligation in a particular turn.

The example below is an excerpt from the transcript obtained from keyboard-to-keyboard session number one. To keep the authenticity of the data we did not correct the typographical mistakes made by the student or tutor. The excerpts in boldface are student initiatives and the excerpts in italics are tutor responses.

Example 1.

K1-tu-26-2: Why would there be an increased amount of blood coming into it?

K1-st-27-1: I guess there would not be increased RAP, but **would the TPR increase because of the increased CO?**

K1-tu-28-1: *Were dealing now with the DR period.*

K1-tu-28-2: *That's before there are any neural, reflex changes.*

K1-tu-28-3: *The TPR is neurally controlled.*

K1-tu-28-4: *So what do you think would happen to it?*

K1-st-29-1: During the DR period it would remain unchanged.

K1-tu-30-1: *Correct.*

The student is unsure whether his explanation is correct and is expressing uncertainty by producing a hedged answer, followed by a question. The identifiers such as “K1-st-27-1” are produced by running a numbering program on the raw transcript. They are used for recording

information such as the fact that this sentence comes from the first keyboard-to-keyboard session, the student's turn, the twenty-seventh turn in the session, and the first sentence in that turn. A turn is a continuous stretch of the dialogue in which where one participant types (the tutor or the student). Each turn consists of one or more (communicative) acts. The identifier "K1-tu-28-1" can be interpreted in a similar way except for one difference; "tu" stands for tutor. Sometimes "tu" and "st" are replaced by "ti" and "si"; these indicate interruptions by the tutor or student respectively.

Our inter-rater reliability study had judges selected from our research group. First, two of the authors, Farhana Shah and Martha Evens, agreed on the classification of all of the student initiatives and tutor responses in the first twenty-eight tutoring sessions. We then started to work with our colleague Stefan Brandle. He read the category descriptions and then we went over the initiatives and responses in sessions one and two together. Next he classified responses and initiatives in sessions three and four, followed by a discussion of each decision. After this orientation he then classified the initiatives and responses in the remaining twenty-four sessions. There are 124 initiatives and responses in the sessions K5-K28.

A Brief Look at Related Literature

Much of the early research in dialogue analysis is based upon the work of Emanuel Schegloff. The definition of turns and turn-taking used here is based on a well-known paper of Schegloff and Sacks (1973). In particular, that landmark paper had a foundational impact on our understanding of the ways in which expert tutors move the conversation back to their own agendas after an initiative.

Maintaining a mixed-initiative dialogue requires the understanding of the intentionality behind an action (Lehnert, 1978). In mixed-initiative dialogue either discourse participant may

take control over the conversation. A transfer of control takes place between the participants (Walker & Whittaker, 1990). For example, in tutoring, students give information that is not requested and pose questions of their own. Participating in a conversation and producing an utterance can be viewed as problems in plan construction. The problem is to create an utterance that satisfies goals involving the transmission of certain contents or intentions (Wilensky, 1983). Traum and Allen (1994) developed a rule-governed approach versus a plan-based approach and proposed that an agent's behavior is a blend of that agent's current goals in the domain and a set of obligations that are shaped by a set of social norms. For example, if an agent is requested to perform an action, that agent is obliged to perform the requested action rather than attaining personal goals. The rules that encode discourse conventions comprise Traum and Allen's model of discourse obligation. The language of the tutorial dialogues that we analyze in this article is shaped by the fact that the tutor's goal is to teach causal reasoning, and, therefore, both tutors and students are attempting to express this kind of reasoning. Papers of Magliano, Baggett, Johnson and Graesser (1993) and Noordman and Vonk (1998) have helped us to understand this causal argumentation better.

Carberry (1990a, b) has played a major role in the growing understanding of the significance of plan recognition. She argues that as intelligent tutoring systems improve their natural language capabilities, "recognition of the student's problem-solving plan will become more and more important (Carberry, 1990b, p. 72)." Carberry's work on handling clarification subdialogues (Lambert & Carberry, 1992) and on understanding indirect speech acts (Green & Carberry, 1994) has motivated our efforts to understand the student's plan. Chu-Carroll and Carberry (1995) model collaborative negotiation as a recursive strategy of squaring inconsistencies. Litman's (1996) analysis of intentions has been particularly important in

classifying communicative goals. Chu-Carroll and Brown (1997) present a strategy for recognizing shifts in initiative that we have tried to adapt to our purposes. The identification of clue words (Moser & Moore, 1995) like “but” and “thought” has helped us in plan recognition. Discourse markers and the underlying coherence relations are very important in carrying on a dialogue (Sanders, Spooren, & Noordman, 1992; Bateman & Rondhuis, 1997). Discourse markers help us to identify the communicative goal of the student. For example, we note that “but” often marks a challenge by the student (Oversteegen, 1997).

The relations between initiative and response adopted in our research differ to some extent from those described by T.J.M. Sanders et al. (1992), because they are dialogue relations, but they are very similar because the tutor's response coheres with the student's initiative in many of the same ways as one sentence of a monologue coheres with earlier sentences. Of course, monologues do not reveal the kinds of conversational repair seen in a dialogue. The relationships we describe here seem to range from the mainly pragmatic (as in conversational repair) to the mainly semantic (as in the examples of teaching the sublanguage) as described by T.J.M. Sanders (1997).

The work closest to ours is Gordon Wells' (1999) *Dialogic Inquiry* which we discovered only after this research project was complete. Wells studied studying educational dialogues between teachers and students and between mothers and children, within a systemic/functional framework. He divided conversational moves into three categories “initiating,” “responding,” and “follow-up.” If we had seen his work earlier we might well have used the term “follow-up” for the question posed by the tutor at the end of almost every turn, even though, in our dialogues, that follow-up is often combined into one sentence with another goal.

Wells' list of "functions" corresponds to our communicative goals. It is longer than ours, partly because his data includes more multiperson dialogues that require explicit turn-taking devices. His list includes many of our goals "request confirmation," "request information," "request repetition" (our "request conversational repair"), "clarify" (our "conversational repair"), and "acknowledge." He divides our "explanation" into "give information" and "give justification /explanation." His "qualify previous contribution" covers much of the same territory as our "teaching the sublanguage."

This paper has three sections. The Methodology section presents our classification of student initiatives and also our classification of tutor responses to those initiatives. The Results and Discussion section describes the relationships we found between initiatives and responses. The Conclusion section summarizes our results and describes some ideas for future work.

METHODOLOGY

In this section we present the categories we developed for student initiatives and then turn to the tutor responses. This is the logical order, because the initiative precedes the response in the actual dialogue. In developing these categories, however, we worked on the tutor responses first, since we could discuss these issues with the actual tutors. The categories for the student initiatives were based on that discussion.

A detailed classification was undertaken by Gregory Sanders several years ago (G. Sanders, Evens, Hume, Rovick & Michael, 1992). He proposed a one dimensional scheme that ending up leaving 35% of the initiatives unclassified. An attempt to establish inter-rater reliability for the scheme failed. Although we abandoned Sanders' categorization scheme, we started with the dialogue markup produced by Sanders and Michael which identified the beginning and ending of each initiative and response. We eventually made a number of changes

in this markup. Following the lead of the DAMSL project (Allen & Core, 1997), we converted our markup to a Standard General Markup Language (SGML) form. This has allowed us to make use of the available SGML utilities for counts of various kinds (McKelvie, Thompson, Tobing, Brew & Mikheev, 1997). We shortened the scope of some of the responses, because Sanders often included many turns and multiple goals in one response. We decided to include pauses as initiatives, since they are not cooperative responses that extend the existing dialogue thread. Sanders included answers like "I don't know" as initiatives, but the expert tutors (JAM & AAR) argued that they find this response helpful and that it should be treated as a cooperative answer. As a result, these answers have been excluded from the list of initiatives. On the other hand, "I don't understand," is treated as an initiative and is classified as a request for information.

We decided to try a multidimensional classification system, since the attempt at a one-dimensional classification scheme had failed. The first two authors (FS & MWE) developed the categorization scheme. The dimensions adopted represented the information needed for automatic recognition of student initiatives and generation of tutor responses. The categories were largely data driven; that is, they were developed during discussions with our expert colleagues, who did the actual tutoring. Except for conversational repair, which seems to occur almost everywhere (Fox, 1990, 1993), the conversational goals we have found are very different from those in the advice-giving and task-oriented dialogues studied by the DAMSL project.

In the interest of producing a classification that could be implemented and readily analyzed statistically, we tried to assign just one communicative goal to each initiative. We were forced to make an exception for acknowledgment and explanation, which appear together in a systematic way. Of course, our experts often combine goals in subtle ways. This

oversimplification was responsible for a number of the differences between the judges in the inter-rater reliability study reported later.

Categorizing Student Initiatives

Student initiatives were categorized along four dimensions: Communicative Goal, Focus or Content, Surface Form, and Degree of Certainty Expressed (i.e., whether the student is hedging or not). The dimensions and types of initiative categories are shown in Table 1. Each of these dimensions represents an axis in a category space that is useful for distinguishing student initiatives. The categories within each dimension are listed in order of frequency in the transcripts. This section introduces our taxonomy of student initiatives in tutoring dialogues.

[Put Table 1 about here]

We describe our categories of student initiatives with examples from the keyboard transcripts of the sessions. Students seem to ask for help in many ways. Both students and tutors use fragments and punctuation marks very frequently. For example, in Example 2 the sentence K5-st-51-1 illustrates the use of I for increase and D for decrease. You will observe also that the student input in K5-st-51-1 is hedged.

Example 2.

K5-tu-50-1: The vascular function curve shows the relationship
 between CO and central venous pressure, when CO is
 the independent variable (i.e. when it changes first).

K5-tu-50-2: Now do you remember?

K5-st-51-1: **So, when CO I, the central venous pressure will D?**

K5-tu-52-1: *Absolutely correct.*

K5-tu-52-2: *What variable is essentially the same as central venous pressure?*

K5-st-53-1: RAP.

K5-tu-54-1: *Right.*

Communicative Goal/Intention. Presumably the student produces illocutionary acts with the intention of fulfilling particular goals. The tutor's understanding of the speech act depends on how he interprets those productions, as a component of what he believes is the student's plan. The different types of student goals are illustrated with examples from the transcripts. Stampe (1975) argued that what makes a request a request is the intention with which it is made; what determines its success is whether its recipient can infer the intention from the linguistic form in context. Agreeing with Stampe's notion we identified a set of goals. Our ultimate plan is to have a computer recognize the goals of the student initiative automatically.

The communicative goals are presented in order of frequency among the 124 initiatives in sessions K5-K28: Request for Confirmation (54), Request for Information (30), Challenge (14), Refusal to Answer (13), and Conversational Repair (13). The same frequencies are given in parentheses in the subheadings and in the bottom row of Table 4.

(1) *Request for Confirmation (N = 54).* The student generates an explanation and asks for confirmation of this theory. Graesser, Person, and Magliano (1995) call this a verification question. Sometimes a simple yes or no is a sufficient response.

Example 3.

K9-st-38-1: **i.e. the change in sympathetic input changes the location of the Starling curve?**

K9-tu-39-1: *Yes.*

More often the tutor responds more elaborately, especially when the student's explanation is wrong as in Example 4.

Example 4.

K10-st-56-2: **Does RAP increase initially with increasing CO and then taper off as CO continues to I?**

K10-tu-57-1: *no. When CO increases it transfers increased quantities of blood from the venous system into the arterial system, decreasing the CBV (central blood volume) and increasing the arterial blood volume (and pressure).*

K10-tu-57-2: *What would happen to the central venous pressure when CBV goes down?*

K10-st-58-1: It decreases.

K10-tu-59-1: *Yes.*

(2) *Request for Information* ($N = 30$). This communicative goal conveys a direct request for information about the topic in focus. Students ask for information in many ways: with yes-no questions as in Example 5, with declaratives as in Example 6, and with wh-questions as in Example 7.

Example 5.

K24-st-35-1: **Did you count my prediction for sv?**

K24-tu-36-1: *Yes, but you haven't predicted tpr.*

Example 6.

K7-st-74-1: **I still don't understand.**

K7-tu-75-1: *Ok. .*

K7-tu-75-2: *Then let's step through it.*

K7-tu-75-3: *What has triggered the RR?*

Example 7.

K13-st-48-1: **If sv d every time hri how can co ever increase**

K13-tu-49-1: *What are the two parameters that determine co?*

K13-st-50-1: Sv and hr

(3) *Challenge* ($N = 14$). This goal reflects a disagreement with what the tutor has said. The use of clue words like *but*, especially at the start of the sentence, often indicates that the initiative is taken as a challenge to the tutor's preceding utterance. The analysis of discourse markers in T.J.M. Sanders et al. (1992) and Sprott (1992) is particularly relevant to the problem of distinguishing the communicative goals in our initiatives. The work of Sanders and Noordman (2000) makes it clear that it is also important from the point of view of student understanding and recall to generate appropriate discourse markers in tutor responses

Example 8.

K20-tu-46-2: But you forgot that the real pacemaker is dead and this guy's HR is determined by the broken artificial pacemaker.

K20-st-47-2: OOPS.

K20-st-47-2: **BUT I WAS JUST READING EARLIER TODAY IN SMITH
AND KAMPINE ABOUT HOW SANS CAN 'TURN ON'
OTHER AREAS AND INFLUENCE HR WITHOUT ACTING
FIRST ON THE SA NODE**

K20-tu-48-1: *It happens sometimes (extopic pacemaker) and sometimes it doesn't.*

K20-tu-48-2: *The description of this patient is asking you to assume that his HR is solely under the control of the artificial pacemaker.*

(4) *Refusal to Answer* ($N = 13$). Sometimes all that the tutor sees is a student pause. The student is busy in working the problem out and needs time to come up with a correct answer or else just gives up. An offer of help from the tutor is the typical response to a student pause.

Example 9.

K7-tu-45-2: An what does that affect?

K7-st-45-a : **{Pause}**

K7-ti-46-1 : *Need help?*

K7-st-47-1: YES!

K7-tu-48-1: RAP is ventricular filling pressure.

Ten of the thirteen cases we classified as refusal to answer are pauses like this one. There are three other situations that we classified as refusal to answer. In one case the student answers "sort of" and in another the student says "I am uncomfortable about my thinking!!!" As a result of our study of pauses, we have redesigned the tutoring system to respond to pauses (defined as no input for a minute) with an offer of help. Of course, the student may be going to the bathroom, getting coffee, making a telephone call or on the way to class, so the system becomes quiescent after fifteen minutes.

(5) *Conversational Repair* ($N = 13$). Both the tutor and the student express their thoughts in a way that is not always perfect or clear. The repair initiative is often a request for clarification.

We include in this category only conversational repair pertaining to miscommunications, not student misconceptions about the domain. Example 10 shows a request for rephrasing.

Example 10.

K4-tu-83-6: How are the falls in TPR and in CC connected to the decrease in
MAP?

K4-st-84-1: **I don't think I understand the question.**

K4-tu-85-1: *What are the determinants of MAP?*

The student asks for repair in turn K4-st-84-1. Basically she asks for restatement of the question in a more precise or specific way. The tutor rephrases the question so that the student can understand it. Next we turn to examining the surface form of student initiatives.

Surface Form. The apparent type of a sentence or phrase can be determined by the word order, mood, and part of speech information. We have included pauses because a long pause on the part of the student causes our expert tutors to drop the current tutoring plan and offer help. The surface form of student input is classified as: Declarative, Interrogative, Silence/Pause, Fragment, or Imperative.

(1) *Declarative (N = 61).* A declarative form makes a statement. Many have requests as an underlying goal in spite of the declarative form, like the student initiative in Example 10 above.

(2) *Interrogative (N = 59).* An interrogative form is a sentence usually recognized as a question by word order or wh-element. Example 5 above contains a student initiative in the form of a yes-no question and Example 7 displays a wh-question initiative. Question marks are not reliable indicators of questions because they are often used as hedges on declarative sentences.

Sometimes, also, students omit the question mark on an actual question. Therefore we classify an initiative as an interrogative only if it has inverted word order or contains a "wh word."

(3) *Silence/Pause* ($N = 11$). Silence/Pause is a form of initiative that should be treated as a special case because of its various nuances. Example 9 above shows a pause. Example 11 shows another example along with the timing information that allows us to recognize a pause. We left the timing information out of the other examples in this paper because it is rather cumbersome to read.

Example 11.

CT= 42:11 ET= 1:26 IT= 0:52

K26-st-67-1: **Is right atrial pressure co×tpr–pa**

K26-tu-68-1: *no. (co ×tpr is map)*

[pause here]

CT= 44:49 ET= 0:45 IT= 0:41

K26-st-69-1: **Need more help**

In this example the symbol CT (Current Time) stands for the time since the session began. ET is the elapsed time in the following turn. IT is the initial response time (between the turn change and the first key stroke). When we first collected the transcripts, it did not occur to us to record timing information, but Fox's (1990, 1993) work convinced us of the importance of this information. A silence is an interruption to an ongoing remark or an occurrence between

remarks. Levelt (1989) defines different types of silence as: pause, gap, or lapse depending on the point of occurrence and duration. In our case, the duration of a pause is at least one minute.

(4) *Fragment* ($N = 9$). These are sentence fragment forms comprising one word or more. Thompson's (1980) empirical study shows that users tend to be as brief as possible. Our transcripts contain a number of initiatives expressed in the form of fragments. In Example 12 a prepositional phrase reveals a major misconception to the tutor:

Example 12.

K26-st-53-1: **By the force lenth relationship?**

K26-tu-54-1: *No!*

K26-tu-54-2: *You are confusing the Frank-Starling effect (increased filling gives greater output) and contractility.*

K26-tu-54-3: *Contractility is the force of contraction AT ANY GIVEN*

FILLING.

(5) *Imperative* ($N = 5$). An imperative form often marks a request. Like the declarative form the imperative sentence is usually followed by a period as in Example 13.

Example 13.

K11-si-18-1: **Let me restate my question.**

Sometimes even an imperative is hedged with a question mark as in Example 14.

Example 14.

K22-st-89-1: **Let me start somewhere else?**

Degree of Certainty – Hedging. Our transcripts of expert tutoring sessions contain many types of hedges in the student input. Almost any speech act can be hedged, although imperatives are not hedged as often as declarative or interrogative sentences in our data. We decided to treat hedging as a separate dimension because it reflects different shades of meaning in the student input. George Lakoff (1972) classified a number of adverbs and adverbial expressions that are used to express hedges.

One reason for pursuing hedging here is the difficulty of distinguishing between hedged answers and initiatives. Example 15 shows a hedged answer to the question posed by the tutor.

Example 15.

K25-tu-52-2: But what determines the volume of blood in the central venous
 compartment?

K25-st-53-1: How about co?

K25-tu-54-1: Certainly, CO is the determinant i'm looking for here.

But students sometimes start off a question with "how about" as in the first line of Example 23 below. How does our tutoring system distinguish between these two? Our current game plan is to try to analyze the student's input as an answer to a question. If this fails, the system assumes that it has encountered an initiative.

Robin Lakoff (1971) addressed the pragmatics of modality and showed that *must*, *cannot*, and other modals function as performatives. The modals *may* and *might* also function as hedges. In our transcripts we see many types of hedges in the form of adverbs like *maybe*, *perhaps*; in the form of verbs like *I guess*; in the form of auxiliary verbs like *may*, *might*, *can not*; in the form of adjectives like *I am not sure*, *I am not comfortable*; in the form of informal expressions like *sort*

of, and most often, one or more question marks ????. The following examples illustrate this variety.

Example 16.

K3-tu-53-1: The venous return may not change for a couple of minutes but what about the rate at which blood is being removed vfrom the central blood compartment?

K3-st-54-1: That rate would increase, **perhaps increasing RAP???**

Example 17.

K15-tu-73-1 Well, if $co = 0$ and $hr = 0$ then sv would have to be 0 and You didn't predict that.

K15-st-74-1 True, I predicted co **because I thought venous return might increase.**

Example 18.

K12-tu-93-1: no. I'm agreeing with you the vessels are dialated.

K12-tu-93-2: I was just giving you information that you could use to determine how the reflex accomplish that.

K12-st-94-1: **Dilation results in increasing the vessel radius and thus tpr goes down (exponentially by a factor of 4???)**

Example 19.

K13-tu-23-1: Now what?

K13-st-24-1: Cc i **maybe**

Example 20.

K13-tu-55-2: One last question here...

K13-tu-55-3: Why did you predict that cc and tpr would be unchanged.

K13-st-56-1: TPR is largely a function of arteriolar constriction.

K13-st-56-2: Cc changes in response to ans stimulation or ca build up during
tachecardia.

K13-st-56-3: **Im not sure if 120bpm is fast enough to cause that**

Example 21.

K24-tu-48-1: Why did you say d?

K24-st-49-1: Because I thought that the pacemaker is stuck at 50, but **I guess
sympathetic come into play here, right?**

Tutor Responses to Student Initiatives

This subsection examines tutor responses to student initiatives in our study. As a reminder, we italicize the tutor responses as opposed to the student initiatives, which are printed in boldface. We feel somewhat greater confidence in this classification because we discussed and confirmed the results with our colleagues and co-authors, the real tutors.

The criteria we adopted fall naturally into three dimensions, as shown in Table 2. These dimensions are the communicative intention, the delivery mode, and the surface form of the expression. The communicative goals determine the choices of what to teach, whereas the delivery modes are associated with how to teach, and the surface forms are related with how to express the message. We were not satisfied with our analysis of the delivery modes, so we decided to eliminate them from this discussion. We feel that the more fine grained analysis by

Hume, Michael, Rovick and Evens (1996), Kim et al. (1998), and Zhou, Freedman, Glass, Michael, Rovick and Evens (1999a, b) is a more satisfactory guide to the choice of delivery mode in generation.

[Put Table 2 about here]

Communicative Goal. “The mother of each speech act is communicative intention.” said Levelt (1989, p.108). Establishment of a communicative goal by the tutor is the first step in the generation of a response. These goals support the introduction of discourse topics that are shared by the tutor and student. We will illustrate these goals one by one. The frequency among the 124 initiatives in sessions K5-K24 is given in parentheses in each subheading. These frequencies also appear in the rightmost column of Table 4, where it is easier to compare them.

(1) *Explanation* ($N = 52$). The response to different requests may be an explanation of varying length. Appropriate explanations are constructed on the basis of the communicative intention and on the content of the student question or theory. Often the explanation targets the underlying parameters and causal relations between the predicted and predictor variables. The tutor varies the mode of delivery; sometimes for the sake of variety and sometimes for a specific purpose. As we have seen, the explanation often ends with a follow up question as in K16-tu-39-4 in Example 22. The student is requesting information to fill the gap. The hedged statement made by the student is an indication of her lack of confidence in her knowledge. She is unsure about the process that controls contractility. The tutor responds with an explanation.

Example 22.

K16-st-38-1: **I think i am getting contractility mixed up with stroke volume...**

K16-st-38-2: **Contractility is the force of contraction that i think goes up with**

increased heart rate, but i am not sure how

K16-tu-39-1: *Ok, let me explain.*

K16-tu-39-2: *The length-tension relationship of muscle says that as length goes up (as the ventricle fills more) the force of contraction will increase.*

K16-tu-39-3: *Changes in contractility result in changes in force at the same fiber length or same filling.*

K16-tu-39-4: *What input to the heart causes contractility to change?*

In Example 23 the student is proposing a mini-theory and asking the tutor for confirmation. The tutor responds with an acknowledgment while launching an explanation. The amount of detail varies.

Example 23.

K4-st-46-1: **Well, if SV is volume pumped per beat, and we already know the number of beats is increased how about the RAP, which may have an effect on how much blood is reaching the ventricle.**

K4-tu-47-1: *Definitely, RAP affects ventricular filling.*

K4-tu-47-2: *What's the relationship?*

K4-st-48-1: **An increased volume means an increased RAP.**

K4-tu-49-1: *I'm not sure I am sure what volume you are talking about in your Previous statement.*

(2) *Acknowledgment* ($N = 14$). An acknowledgment tells the student whether his/her explanation was understood, and usually, whether it is correct or not. It provides evidence of

understanding in response to the student initiative. The acknowledgment may be positive or negative depending upon the content presented by the student. It may be made in a very simple manner, as in Example 24, or a more or a rather complex manner, as in Example 25.

Example 24.

K6-st-60-1: **Does the direct affect steady state more than the reflexes?**

K6-tu-61-1: *Yes.*

Example 25 illustrates a compound tutor response that consists of a negative acknowledgment, followed by an elaboration. The tutor usually generates an explanation after a negative acknowledgment. We classify examples with an acknowledgment followed by an explanation as explanations.

Example 25.

K12-tu-33-1: By what mechanism will it increase?

K12-st-34-1: **If you increase pressure will you momentarily increase resistance**

K12-tu-35-1: *No.*

K12-tu-35-2: *You may be thinking of autoregulation.*

K12-tu-35-3: *That's slow.*

K12-tu-35-4: *Remember that we're dealing with the short period before you get a reflex response.*

(3) *Conversational Repair* ($N = 14$). Repair is done to avoid misunderstanding and to correct misconceptions (McRoy & Hirst, 1995). Fox argues (1993) that such repairs involve, in effect, a reconstruction of the initial utterance. If the misunderstanding is not noticed at once, the conversation may break down at a later stage. So it is very important to make an attempt to

resolve the issue immediately. Presently we are restricting ourselves to conversational repair of miscommunication; we exclude repairs of misconceptions in cardiovascular physiology. Examples 26 and 27 depict two common forms of conversational repair.

Example 26.

K5-tu-83-2: Now, where do you want to go?

K5-st-84-1: CO D.

K5-tu-85-1: Right.

K5-tu-85-2: Why?

K5-st-86-1: Because CC D.

K5-tu-87-1: So?

K5-st-88-1: **I don't understand.**

K5-tu-89-1: *How does CC D affect CO?*

Example 27.

K2-st-9-1: **What do you mean by "pass the effect"?**

K2-tu-10-1: *I mean that the system behaves as though the left ventricle was filling from the right atrium.*

K2-st-11-1: Ok

(4) *Instruction in the "Rules of the Game"* ($N = 12$). In this response the tutor is telling the student how to proceed. The student has forgotten the protocol or does not understand what the tutor expects.

Example 28.

K27-st-13-1: **Predict using arrows?**

K27-tu-14-1: *You can fill out the prediction table i gave you using arrows or +/-0.*

(5) *Extending Help in Response to Pause (N = 10)*. When the tutor notices a delay on the student side, he intervenes to offer his help. The label on the last line of Example 29 contains “ti” instead of “tu” to indicate that this is a tutor interruption.

Example 29.

K5-st-42-1: Yes, $MAP=CO * TPR$.

K5-tu-43-1: Right.

K5-tu-43-3: How about yet another variable.

K5-tu-43-2: We went through that before.

K5-st-45-1: **I don** [big pause here]

K5-ti-46-1: *Need help?*

(6) *Teaching the Problem Solving Algorithm (N = 10)*. A major goal of the tutor is making sure that the student understands how to solve problems. The student needs to learn how to carry out a sequence of steps using causal reasoning. In the example below the tutor responds to the initiative by prodding the student to start at the right point.

Example 30.

K12-tu-61-6: What I want to know is how the reflex is going to get SV to D?

K12-st-62-1: Decrease filling time, decrease venous return.

K12-st-62-2: **I'm just hesitant to say what comes first.**

K12-st-62-3: **I'll go with tpr i to slow blood flow back to heart (i don't really like this idea)**

K12-tu-63-1: *Well let's see if we can get at the first question I asked and then we'll come back to TPR.*

(7) *Probing the Student's Inference Processes* ($N = 5$). The tutor encourages the student in active learning through self explanation (Chi et al. 1994). This provides the tutor with feedback that helps him in student plan recognition. It also helps the tutor update his model of the student and generate justifications of his own reasoning and behavior. For example:

Example 31.

K5-st-102-2: **But I'll bet that's not right.**

K5-tu-103-1: *Well you're right in your bet.*

K5-tu-103-2: *SV D because CC D.*

K5-tu-103-3: *That doesn't mean that RAP has to be D!*

K5-tu-103-4: *Let me remind you again of the vascular function curve.*

K5-tu-103-5: *Does that help?*

K5-st-104-1: **RAP I.**

K5-tu-105-1: *Would you explain.*

K5-tu-105-2: *You're right but I just want to hear what you'r thinking.*

(8) *Brushing Off* ($N = 5$). Sometimes the tutors decide to avoid or put off a discussion and bring the dialogue back to issues of higher priority. This kind of response is used most often

when the tutors do not understand what the student is driving at. In the following example the tutor seems to take control of the dialogue, pursuing his own agenda.

This is one area where our analysis of the transcripts has influenced the behavior of the tutors. In the first eight sessions, the tutors almost never brushed off initiatives, but when we analyzed examples of repair in those sessions (following Fox, 1990) we showed that misunderstanding of a student initiative was second only to ambiguous "how" questions as a cause of conversational repair. Our expert tutors were horrified by these examples; they stopped asking "how" questions and started brushing off student initiatives when they were not sure what the student wanted.

Example 32.

K15-st-82-1: Dr 0, rr d, therefore ss d

K15-tu-83-1: Good thinking!

K15-tu-83-2: What next?

K15-st-84-1: Cc d, but **i really still dont understand why**

K15-tu-85-1: *You're right and we'll talk about it further if there is time.*

(9) *Teaching the Sublanguage (N = 2).* The tutor is concerned about teaching the correct usage of the language of physiology. Indeed, this is one of the most important reasons for implementing a natural language dialogue in CIRCSIM-Tutor. We believe that the diagnosis and remediation of student problems with language use requires a natural language interface. Example 33 gives a good illustration.

Example 33.

K12-st-46-1: **Does the rate of blood removal from the central veins mean that blood entering the right atrium, if so ithink venous return does**

go up immed.

K12-tu-47-1: *We need to get our terminology straight.*

K12-tu-47-2: *Venous return means blood returning from the systemic circulation to the heart.*

Surface Form. Since our goal is to recognize student initiatives and generate responses automatically, we hoped to discover that the goal of the initiative predicts the surface form. However, if there is a relationship it is not a simple one. If we re-examine the examples of student requests for conversational repair already shown in this paper, we see a variety of surface forms:

Example 10. K4-st-84-1: **I don't think I understand the question.**

Example 9. K7-st-45-a: {**pause**}

Example 13. K11-si-18-1: **Let me restate my question.**

Example 26. K5-st-88-1: **I don't understand.**

Example 27. K2-st-9-1: **what do you mean by "pass the effect"?**

When it comes to tutor responses, we see the same kind of variety. Acknowledgments are often expressed in fragments but sometimes they are conveyed by a declarative sentence as in:

Example 6. K20-tu-48-1: *It happens sometimes (extopic pacemaker) and sometimes it doesn't.*

Example 12. K26-tu-54-2: *You are confusing the Frank-Starling effect (increased filling gives greater output) and contractility.*

Explanations are often expressed in declarative sentences or sequences of declarative sentences, but we see a great deal of variety here too. Example 22 contains a relatively long tutor

explanation. It begins with an imperative (K16-tu-39-1), and winds up with a question (K16-tu-39-4).

Sometimes the explanation and the question get combined into one, as in Example 2 K5-tu-52-2: "What variables are essentially the same as central venous pressure?" The response to a request for repair can be just a restated question as in Example 8 K4-tu-85-1: "What are the determinants of MAP?"

We suspect that the choice of surface form in the tutor response is determined by all the dimensions combined and also by the need to keep the dialogue going. As a result, most tutor responses end in a question, but sometimes the question can be implicit in a declarative sentence, as we see in K5-tu-105-2 from Example 31 "You're right but I just want to hear what you'r thinking." We do believe that communicative goals and their context determine discourse markers and we have derived rules (using a machine learning program) that produce appropriate discourse markers (Kim et al., 2000).

RESULTS AND DISCUSSION

In this section we present several results that are relevant to our plans for implementation. First, we show that student initiatives are frequent enough to be important for our system to support. Second, we present our inter-rater reliability studies, which make a convincing case for the validity of our categories. Third, we describe some significant relationships between student initiative categories and tutor response categories. Therefore, if our system can recognize the initiatives we can generate appropriate responses.

We identified and classified 146 student initiatives in 28 hour-long human tutoring sessions. Table 3 represents data for the 28 transcripts. Men and women produced the same number of initiatives per session on the average (5.1 vs. 5.4), but the women students are more likely to hedge (47% for the women vs. 27% for the men) with $\chi^2 = 5.90$ ($p < 0.5$).

These 28 sessions contain 2666 turns altogether, or approximately 1333 student turns. Thus, on the average, there was one initiative for every nine turns. However, the initiatives were not evenly distributed. Two of the 28 students produced no initiatives at all, while, at the other end of the spectrum, 2 students produced 11 initiatives and 1 produced 12. Both the tutors (our collaborators, Joel Michael and Allen Rovick) were male. The students, all first year medical students and paid volunteers, included 9 men and 19 women. All the sessions involved problems with a pacemaker; in most the heart rate increased (HR I in Table 3) in others the heart rate decreased (HR D in Table 3). This difference is apparently unrelated to the other variables in the table.

[Put Table 3 about here]

The results of our classification of student initiatives and tutor responses are shown in Table 4. Our colleagues (JAM & AAR) believed that the perceived student goal is the most important factor in choosing a communicative goal for the response. This motivated us to combine both sets of goals in one table.

Each category of student initiative is represented by a column in Table 4 and the total frequencies for these categories appear across the bottom of the table. These results show that the most frequent category is request for confirmation (N=54). In these initiatives the student is presenting a theory and asking the tutor to verify it. This seems to be the conversational reflection of the self-explanation process that Chi et al. (1994) have shown to be so crucial to

effective learning. The next most frequent category is request for information (N=30); the student is actively asking the tutor for an explanation. Graesser and his students (Graesser 1992; Graesser & Person, 1994) have demonstrated the important place that question-asking holds in learning new concepts. The last three categories are all relatively close in size. The challenge category (N=14) may perhaps represent another aspect of self-explanation, in which the student asks what is wrong with a previously held theory. Refusal to answer is in fourth place (N=13). We presume that this lack of response will be less frequent with the machine tutor because the students should be less concerned about sounding stupid to our program than with a professor. Last comes conversational repair (also, N=13). Our colleagues consider these examples to result from errors on their part. They tell us to make sure that CIRCSIM-Tutor never asks an ambiguous question. Of course, we cannot hope to implement the whole range of responses that human tutors produce. As Freedman (1997) has pointed out we can only hope to implement responses that we can integrate with our own system goals.

In the inter rater reliability study of student initiative categories described earlier, we mentioned that sessions K1-K4 were used for training so we report Cohen's Kappa values for K5-K28 (Carletta, 1996). The 2x2 Cohen's Kappa reliability scores were .83, .96, .71, .82 and .78 for request for confirmation, request for information, conversational repair, challenge, and refusal to answer, respectively. For the combined classification table, we obtained $K = 0.745$, so we feel that our classification of the Student's Communicative Goals is reliable.

The results of our study of tutor responses can also be seen in Table 4, where the total frequencies for each type of tutor response run down the right side. The most frequent category of tutor response is explanation (52). The other categories appear less often: Acknowledgment alone (14), Conversational Repair (14), Instruction in the Rules of the Game (12), Teaching the

Problem-Solving Algorithm (10), Help in Response to Pause (10), Probing the Student Inference Process (5), Brush-off (5), and Teaching the Sublanguage (2).

Our inter rater reliability study of tutor responses showed 2x2 Cohen's Kappa scores of .82 for Explanation, .91 for Acknowledgment, .78 for Conversational Repair, .82 for Instruction in the Rules of the Game, .76 for Teaching the Problem-Solving Algorithm, .48 for Probing the Student's Inference Process, 1.0 for Help in Response to Pause, .88 for Brush-Off, and .663 for Teaching the Sublanguage. For the combined classification table we obtained an overall Kappa of .832, so we feel that our classification of the Communicative Goals of the tutor responses to student initiatives is reliable. Examination of the responses classified as Probing the Student's Inference Process suggested that this goal was often combined with others and that the raters might have agreed if we had allowed them to assign two goals to a single response.

Relationships between initiatives and responses are particularly important to our plans for implementation. If the student has requested confirmation of a correct theory, the tutor often stops with an acknowledgment. If the theory is wrong an explanation is required. Challenges also require an explanation. The next most frequent category of tutor response is conversational repair, almost always, as you might expect, in response to a request for repair. The tutors explicitly ask if the student wants help when the student refuses to answer. Though they often provide help at other times, they never ask this kind of question except after a pause. At other times, they offer hints, usually embedded in questions, and explanations. They teach the problem-solving algorithm (TPSA, N=10) when they detect a confusion about the order of predictions. They explain the rules of the game (IROG, N=12) when the student seems to be confused about the meaning of the stages. If we split the data in Table 4 up into 2x2 transition tables we find that some of these relationships are significant. Using Fisher's exact test we get p

< .001 for the relationship between request for repair and conversational repair. A request for confirmation is generally followed by an acknowledgment with an accompanying explanation ($\chi^2 = 27.48$). Challenge typically gets the same response. There is also a significant relationship between a request for information and a response of explanation ($\chi^2 = 11.53$). The other combinations did not give significant results using the χ^2 test.

The analysis of the surface form of student initiatives confirmed our suspicion that requests are not marked by question marks or question forms. Indeed, question marks are at least as likely to indicate hedged answers. Counting the sentence forms gave 61 declarative sentences, 59 interrogatives, 9 fragments, and 5 imperatives. In addition there were 11 pauses. The categories of surface form turned out to be easy for coders to understand and we obtained a Cohen's Kappa value over 95%.

Our study of the focus of attention or content showed that this area was easy to code and we achieved Kappa over 90%. We will not report the results in detail because we suspect that they will not generalize to other domains. We will not report the results of our analysis of delivery mode either, because they were inconclusive. We think that the categories used here were too broad; as a result they overlap. We suspect that the delivery mode is not really independent of the other dimensions but determined by the content and the communicative goal, as we have found in dialogues where the tutor retains the initiative (Hume et al., 1996; Freedman, Zhou, Kim, Glass & Evens, 1998; Kim et al., 1998).

The frequency counts for different kinds of surface form in tutor responses are: declaratives (197), interrogatives (101), fragments (100), and imperatives (39). Since the responses sometimes involved several sentences, the total is larger than 146. Again the surface form was easy to code and we obtained a Kappa score above 95%. The tutors do not pause and

we assume that the machine tutor should not pause either. Human tutors use imperatives more than students do. Perhaps, this has something to do with perceived status. The machine tutor will use imperatives only rarely.

CONCLUSION

This paper has described our classification schemes for student initiatives and tutor responses. We used four dimensions for classifying the student initiatives: Communicative Goal, Focus/Content, Surface Form, and the Degree of Certainty. We used three dimensions in classifying the tutor responses: Communicative Goal, Delivery Mode, and Surface Form. The tutors normally set a goal and devised a strategy to accomplish it in an effective manner. In this process new subgoals were set up and this process repeated itself. Men and women produced roughly the same number of initiatives per session on the average, but the women students were more likely to hedge.

Our results support our plans to implement responses to student initiatives in a machine-generated tutoring dialogue. They give us a basis for recognizing initiatives and producing appropriate responses. The most common category of student initiatives was a conversational reflection of the self-explanation process, which has been shown to be a critical part of learning. A challenge is apparently a more confrontational version of the same process. Requests for information and the explanations elicited in response show the student in the process of active questioning, also an important part of active learning. The relationships we have found between the student goal and the tutor goal give us a basis for planning responses. The study of sentence forms and hedging suggests strategies to distinguish between initiatives and hedged answers.

We have measured the inter-rater reliability of our classification scheme by using Cohen's Kappa coefficient to examine the agreement between pairs of judges. The level of agreement for trained judges was quite high and provides valuable support for our classification.

An important long-term goal of this research is to build a computational model of tutor responses to student initiatives for our intelligent tutoring system, CIRCSIM-Tutor. Another goal, of course, is to increase our scientific understanding of tutoring dialogues. We have taken a preliminary look at initiatives in novice tutoring sessions in the same domain. It is immediately evident that novice tutors are much more likely to tell students the answers, whereas experts elicit the answers from the students (Glass et al., 1999). Expert tutors also elicit more initiatives than novices do. This evidence further supports our conviction that student initiatives play an important role in tutoring discourse.

Author Notes: This work was supported by the Cognitive Science Program, Office of Naval Research, under Grant No. N00014-94-1-0338, to Illinois Institute of Technology. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

We also wish to thank the University of Quaid-I-Azam in Islamabad for major support of Dr. Farhana Shah.

A much earlier draft of this paper was presented at the 1997 American Association for Artificial Intelligence (AAAI) Spring Symposium on Computational Initiatives in Mixed Initiative Interaction (Shah & Evens, 1997). The reviewers and participants provided many helpful suggestions on our research. The anonymous reviewers of *Discourse Processes* have also helped us to clarify our presentation.

Anyone who wants a copy of CDS (the Computer Dialogue System) or of CIRCSIM-Tutor should notify Martha Evens (evens@iit.edu).

REFERENCES

- Allen, J., & Core, M. (1997). Draft of DAMSL: Dialog act markup in several layers. Available from the University of Rochester, Department of Computer Science at <ftp://ftp.cs.rochester.edu/pub/packages/dialog-annotation/manual.ps.gz>
- Bateman, J.A., & Rondhuis, K.J. (1997). Coherence relations: Toward a general specification. *Discourse Processes*, 24 (1), 3-50.
- Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.
- Carberry, S. (1990a). Incorporating default inferences into plan recognition. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 471-478). Cambridge, MA: MIT Press.
- Carberry, S. (1990b). *Plan recognition in natural language dialog*. Cambridge, MA: MIT Press.
- Carletta, J. (1996). Squibs and discussions, Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22 (2), 249-254.
- Chi, M., Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439-477.

- Chu-Carroll, J., & Carberry, S. (1995). Response generation in collaborative negotiation. 33rd *Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA, 136-143.
- Chu-Carroll, J., & Brown, M. (1997). Initiative in collaborative interactions – Its cues and effects. Technical Report SS-97-04, *American Association for Artificial Intelligence (AAAI) Spring Symposium*, (pp. 16-22), Stanford, CA, 16-22.
- Fox, B. (1990). *Final report on the human tutoring dialogue project*, Technical Report. Department of Linguistics, University of Colorado, Boulder, Colorado.
- Fox, B. (1993). *The human tutorial dialogue project: Issues in the design of instructional systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Freedman, R. (1997). Degrees of mixed-initiative interaction in an intelligent tutoring system. *American Association for Artificial Intelligence 1997 Spring Symposium: Computational Models for Mixed Initiative Interaction*. (pp. 44-49) Stanford, CA.
- Freedman, R., Zhou, Y., Kim, J., Glass, M., & Evens, M.W. (1998). SGML-based markup as a step toward improving knowledge acquisition for text generation. *American Association for Artificial Intelligence 1998 Spring Symposium: Applying Machine Learning to Discourse Processing*, Stanford, CA.
- Glass, M., Kim, J.H., Evens, M.W., Michael, J.A., & Rovick, A.A. (1999). Novice vs. expert tutors: A comparison of style. *Proceedings of Midwest Artificial Intelligence and Cognitive Science Conference 99* (pp. 43-49). Bloomington, IN.
- Graesser, A.C. (1992). *Questioning mechanisms during complex learning*. Memphis State University, Memphis, TGN (ERIC Document Reproduction Service ED No. 350 306).

- Graesser, A.C., & Person, N.K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 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, 495-522.
- Green, N., & Carberry, S. (1994). Generating indirect answers to yes-no questions. *Proceedings of Seventh International Workshop on Natural Language Generation* (pp. 189-198). Kennebunkport, ME.
- Green, N., & Carberry, S. (1999). A computational mechanism for initiative in answer generation. *User Modeling and User-Adapted Interaction*, 9(1&2), 93-132.
- Grice, H.P. (1975). Logic and conversation. In P. Cole and J. Morgan (eds.) *Syntax and Semantics*. (pp, 41-58). New York City, NY: Academic Press.
- Hume, G., Michael, J., Rovick, A., & Evens, M.W. (1996). Hinting as a tactic in one-on-one tutoring. *Learning Sciences* 5 (1), 23-47.
- Kim, J.H., Freedman, R., & Evens, M.W. (1998). Relationship between tutorial goals and sentence structure in a corpus of tutoring transcripts. *Ninth Midwest Artificial Intelligence and Cognitive Science Conference* (pp. 124-131). Dayton, OH.
- Kim, J., Glass, M., Freedman, R., & Evens, M.W. (2000). Learning the use of discourse markers in tutorial dialogue for an intelligent tutoring system. *Proceedings of Cognitive Science* (pp. 262-267). Philadelphia, PA.
- Lakoff, G. (1972). Hedges: A study in measuring criteria and the logic of fuzzy concepts. *Eighth Regional Meeting of the Chicago Linguistic Society* (pp. 183-228). Chicago, IL: University of Chicago.

- Lakoff, R. (1971). If's, and's, and but's about conjunction. In C. Fillmore & D. Langendoen (eds.), *Studies in Linguistic Semantics* (pp. 114-178). Chicago, IL: Holt, Reinhart and Winston.
- Lambert, L., & Carberry, S. (1992). Modeling negotiation subdialogues. *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics* (pp. 193-200). Newark, DE.
- Lehnert, W. (1978). *The process of question answering: A computer simulation of cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Li, J., Seu, J., Evens, M.W., Michael, J., & Rovick, A. (1992). Computer dialogue system (CDS): A system for capturing computer-mediated dialogue. *Behavior Research Methods, Instruments, and Computers*, 24 (4), 535-540.
- Litman, D. (1996). Cue phrase classification using machine learning. *Journal of Artificial Intelligence Research*, 5, 53-94.
- Magliano, J.P., Baggett, W.B., Johnson, B.K., & Graesser, A.C. (1993). The time course of generating causal antecedent and causal consequent inferences. *Discourse Processes*, 16 (1), 35-53.
- McKelvie, D., Thompson, H., Tobing, R., Brew, C., & Mikheev, A. (1997). The normalized SGML library LT NSL version 1.5. Available from: <http://www.ltg.ed.ac.uk>. Edinburgh: Language Technology Group, Human Communication Research Centre, University of Edinburgh.
- McRoy, S. & Hirst, G. (1995). The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21 (4), 435-478.

- Michael, J.A., Rovick, A.A., Evens, M.W., Shim, L., Woo, C. & Kim, N. (1992). The uses of multiple student inputs in modeling and lesson planning in CAI and ICAI programs. *Proceedings of ICCAL Conference*, I. Tomek (ed.) (pp. 441-452). Wolfville, Nova Scotia.
- Michael, J.A., & Rovick, A.A. (1993). Comparing human tutoring with a selected reading assignment. Unpublished paper. Department of Physiology, Rush Medical College, Chicago, IL.
- Moser, M., & Moore, J.D. (1995). Investigating cue selection and placement in tutorial discourse. In *33rd Annual Meeting of the Association for Computational Linguistics* (pp. 130-135). Cambridge, MA.
- Noordman, L.G.M., & Vonk, W. (1998). Memory-based processing in understanding causal information. *Discourse Processes*, 26 (2&3), 191-212.
- Oversteegen, L. (1997). On the pragmatic nature of causal and contrastive connectives. *Discourse Processes*, 24 (1), 51-86.
- 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.
- Sanders, G., Evens, M., Hume, G., Rovick, A., & Michael, J. (1992). An analysis of how students take the initiative in keyboard-to-keyboard tutorial dialogues in a fixed domain. In *14th Annual Meeting of the Cognitive Science Society* (pp. 1086-1091). Bloomington, IN.
- Sanders, T.J.M., Spooren, W.M., & Noordman, L.G.M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15 (1), 1-37.

- Sanders, T.J.M. (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, 24 (1), 119-148.
- Sanders, T.J.M., & Noordman, L.G.M. (2001). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60.
- Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8 (3), 289-327.
- Shah, F. & Evens, M.W. (1997). Student initiatives and tutor responses in a medical tutoring system. Technical Report SS-97-04, *American Association for Artificial Intelligence Spring Symposium on Computational Models for Mixed Initiative Interaction*, Stanford, CA.
- Sprott, R. (1992). Children's use of discourse markers in disputes: Form-function relations and discourse in child language. *Discourse Processes*, 15 (4), 423-439.
- Stampe, D. (1975). Meaning and truth in the theory of speech acts. In Peter Cole & Jerry L. Morgan (eds.): *Speech acts: Syntax and semantics* (pp. 1-40). New York, NY: Academic Press.
- Thompson, B. (1980). Linguistic analysis of natural language communication with computers. *Proceedings of COLING 80*, Tokyo, Japan. np.
- Traum, D., & Allen. J. (1994). Discourse obligation in dialogue processing. *32nd Annual Meeting of the Association for Computational Linguistics* (pp. 1-8). Las Cruces, NM.
- Walker, M., & Whittaker, S. (1990). Mixed initiative in dialogue: An investigation into discourse segmentation. *28th Annual Meeting of the Association for Computational Linguistics* (pp. 77-78). Philadelphia, PA.
- Wells, G. (1999). *Dialogic inquiry*. Cambridge UK: Cambridge University Press.

- Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning*. Reading, MA: Addison-Wesley.
- Woo, C.W., Evens, M.W., Michael, J.A., & Rovick, A.A. (1991). Dynamic planning in an intelligent cardiovascular tutoring system. *Fourth Annual IEEE Symposium on Computer Based Medical Systems* (pp. 226-233). Baltimore, MD.
- Zhou, Y., Freedman, R., Glass, M., Michael, J.A., Rovick, A.A., & Evens, M. 1999a. What should the tutor do when the student cannot answer a question? *Proceedings of the Florida Artificial Intelligence Research Symposium* (pp. 187-191). Orlando, FL.
- Zhou, Y., Freedman, R., Glass, M., Michael, J.A., Rovick, A.A., & Evens, M. 1999b. Delivering hints in a dialogue-based ITS. *Proceedings of the American Association for Artificial Intelligence* (pp. 128-134). Orlando, FL.

Table 1**Classification of Student Initiatives in Four Dimensional Space****Dimension 1: Communicative Goal****Request for Confirmation****Request for Information****Challenge****Refusal to Answer****Conversational Repair****Dimension 3: Focus****Language Issue****Causal Reasoning****Problem-Solving Algorithm****Rules of the Game****Dimension 2: Surface Form****Declarative****Interrogative****Pause/Silence****Fragment****Imperative****Dimension 4: Hedging****Hedged****Not Hedged**

Table 2

Classification of Tutor Responses in Three Dimensional Space

Dimension 1: Communicative Goal

Explanation

Acknowledgement

Conversational Repair

Instruction in the “Rules of the Game”

Help in Response to Pause

Teaching the Problem Solving Algorithm

Probing the Student's Inference Process

Brushing Off

Teaching the Sublanguage

Dimension 3: Delivery Mode

Hinting

Directed Line of Reasoning

Monologue

Rephrasing

Analogy

Dimension 2: Surface Form

Declarative

Interrogative

Fragment

Imperative

Table 3**Data for Twenty-Eight Sessions of Keyboard-to-Keyboard Tutoring**

Session Number	Tutor	Student Age	Student Gender	Proc.	# of Turns	# of Initiatives	Unhedged Initiative	Hedged Initiative
1	AR	22	F	HR I	80	8	2	6
2	AR	28	F	HR I	51	6	3	3
3	AR	23	F	HR I	82	1	0	1
4	AR	25	F	HR I	92	7	5	2
5	AR	22	M	HR I	113	7	2	5
6	AR	22	F	HR I	69	4	4	0
7	AR	23	F	HR I	115	8	2	6
8	AR	25	F	HR I	80	2	0	2
9	AR	22	M	HR I	39	8	6	2
10	AR	27	F	HR I	67	6	6	0
11	AR	26	F	HR I	88	7	7	0

12	JM	25	F	HR I	101	12	2	10
13	JM	27	F	HR I	83	7	4	3
14	JM	27	F	HR I	103	2	2	0
15	JM	34	M	HR I	96	2	1	1
16	JM	26	F	HR I	84	7	5	2
17	AR	22	M	HR D	69	4	3	1
18	AR	27	F	HR D	78	2	1	1
19	AR	26	F	HR D	92	0	0	0
20	JM	25	F	HR D	100	11	5	6
21	JM	27	F	HR D	82	0	0	0
22	JM	27	F	HR D	120	5	2	3
23	JM	34	M	HR D	83	0	0	0
24	JM	26	F	HR D	101	2	1	1
25	JM	32	M	HR I	165	11	9	2
26	JM	24	M	HR D	155	5	4	1
27	JM	23	M	HR D	167	8	7	1

28	AR	23	M	HR D	111	4	4	0
					2666	146	87	59

Table 4**A Transition Table for Student Initiatives and Tutor Responses in Sessions K5-K28**

Initiative Response	Request for Confirmation	Request for Information	Challenge	Refusal to Answer	Repair	Total
Explanation	25	16	11	0	0	52
Acknowledgment	9	5	0	0	0	14
Conversational Repair	2	0	0	1	11	14
Instruction in the Rules of the Game	8	3	1	0	0	12
Teaching Problem Solving Algorithm	5	3	1	0	1	10
Help in Response To Pause	0	0	0	10	0	10
Probing the Student Inference Process	2	1	0	2	0	5
Brush off	1	2	1	0	1	5
Teaching the Sublanguage	2	0	0	0	0	2

Total	54	30	14	13	13	124
-------	----	----	----	----	----	-----
