

# **An Analysis of How Students Take the Initiative in Keyboard-to-Keyboard Tutorial Dialogues in a Fixed Domain**

**Gregory A. Sanders, Martha W. Evens, and Gregory D. Hume**

Department of Computer Science  
Illinois Institute of Technology  
10 West 31st Street  
Chicago, IL 60616  
(312) 567-5153  
sanders@iitmax.iit.edu  
csevens@minna.iit.edu  
ghume@valpo.bitnet

**Allen A. Rovick and Joel A. Michael**

Department of Physiology  
Rush Medical College  
Rush Presbyterian-St. Luke's Medical Center  
Chicago, IL 60612

## **Abstract**

By student initiatives we mean productions which the student could reasonably expect to modify the course of the tutorial dialogue. Asking a question is one kind of student initiative. This paper describes a system called CircSim-Tutor which we are building, the background of the project, the 28 hour-long tutoring sessions analyzed in this paper, and the analysis done. It compares our work to previous work, gives a classification of the student initiatives found and of the tutor's responses to them, and discusses some examples.

---

This work was supported by the Cognitive Science Program, Office of Naval Research under Grant No. N00014-89-J-1952, Grant Authority Identification Number NR4422554, to Illinois Institute of Technology. The content does not reflect the position or policy of the government, and no official endorsement should be inferred.

## **Background of Project**

We are building an Intelligent Tutoring System, called CircSim-Tutor, to tutor first-year medical students on how the body maintains a stable blood pressure from minute to minute, compensating for any perturbation of the pressure. The physiological processes involved are an example of a negative-feedback control system. Such control systems also occur in electronic and mechanical systems that have no connection with physiology. CircSim-Tutor communicates with students in English: it both understands and produces English. As background for this effort, we collected and analyzed 28 sessions, each approximately an hour long, with expert human tutors, Physiology professors at Rush Medical College, doing keyboard-to-keyboard tutoring of first-year medical students attending Rush Medical College, who had heard the class lectures about the material. During these sessions, the student was in one room and the tutor in another room, communicating only by typing on the keyboard and reading from the screen of a computer terminal. Each student always knew that the interaction was with a human tutor and also knew the identity of the tutor. In addition, each student also had the tutor as his or her professor in the related physiology course, being

taken at the same time as part of his or her first-year medical school coursework at Rush. Thus, the students may have felt the academic or social pressure this would imply. All students were volunteers, recruited from these classes by the tutors, and each was paid a nominal amount for his or her participation. A total of 20 students are represented in these 28 sessions (8 students appear twice).

The two Physiology Professors at Rush who are the tutors in these sessions have taught the related physiology courses to first-year medical students at Rush for many years, and have customarily tutored some students taking these classes face-to-face. Thus, our tutors are highly experienced and expert at teaching the material, both in classroom lecture and in personal tutoring sessions.

Transcripts including timing information were automatically collected. Each session was organized as a clinical problem, where a mechanical heart pacemaker suddenly failed, increasing the heart rate (beats/minute) substantially. The student was asked to predict the direction of change, if any, of seven basic cardiovascular parameters, first for the immediate physical effects of the increased heart rate, then for the reflex compensation by the autonomic nervous system to return the blood pressure toward the original value, and finally for the steady state result after this compensation is complete. In addition to making the correct predictions, the tutors want the students to be able to explain why and how each of the changes occurs, and to do so using the "correct" language. It is this concern with language that initially prompted the entire project of building CircSim-Tutor. CircSim-Tutor is a joint project of the Physiology department at Rush Medical College and the Illinois Institute of Technology Computer Science department.

In trying to make CircSim-Tutor handle the discourse phenomena in the sessions with human tutors, we set out to analyze the transcripts of the 28 keyboard-to-keyboard sessions, to identify and categorize each instance where the student took the initiative and to describe how the tutor responded to the initiative. Although the initial purpose of this discourse analysis was to enable the program to respond to such initiatives, we soon became interested in this analysis in its own right.

## Related Work

Graesser, Lang, and Horgan (1988) proposed an analytic scheme for questions, covering a corpus of

approximately 1000 questions asked by adults in different discourse contexts. They proposed 12 semantic categories for questions.

Verification: Is X true or false?

Disjunctive: Is X or Y the case?

Concept completion: Who? What? When? Where?

Feature specification: What is the value of a variable?

Quantification: How much? How many?

Causal antecedent: What caused some event to occur?

Causal consequence: What happened as a consequence of X occurring?

Goal orientation: Why did an agent do some action?

Enablement: What is needed for an agent to do some action?

Instrumental/procedural: How did the agent perform an action?

Expectational: Why isn't X occurring?

Judgemental: What should an agent do?

We seem to need an added category: questions about ontology or taxonomy.

Graesser et al. also proposed 6 pragmatic categories, intended to be orthogonal to the semantic categories. These categories are: information acquisition, assertions, establishing a context for subsequent discourse, indirect requests for non-verbal behavior, conversation monitoring, and humor. While these may cover the questions we found, it is not clear just where they cover repair questions (e.g., "What did you mean?"), investigated by Fox (1990). Nor is it clear to us just where they cover questions intended to establish the relevance of certain facts or cases to the current discourse focus, so as to enable the dialogue to go forward with necessary shared context.

The following is our classification of the student initiatives. Although primarily semantic or pragmatic (generally, discourse-structure based), some of the categories pick out surface clues that seem to flag a production as an initiative. There are 32 sub-categories in the following table, grouped into 12 major categories. This classification was created from study of the 28 sessions.

**Student asks a question.**

Straight question about physiology/physics -- about locally current discourse context

Straight question about physiology -- not about locally current context

**Student makes a Physiology statement (perhaps incorrect)**

Physiology statement -- not in an "answer context"

In response to being asked to make a **corrected** prediction, the student makes some (perhaps accurate) statement of physiology.

I'm not sure if <stmt>.

A "complex" statement, hedged by a '?' (not just, e.g., "Up?" or "CC?")

Maybe I should clarify <previous stmt(s)>

**Student having trouble "seeing" <X>**

I am having trouble seeing/conceptualizing/grasping ...

I am still unclear about <something just discussed>

I think I am getting <X> mixed up with <Y>.

Tutor: "Understand?" Student: "No"

**Student requesting Repair (student does not understand)**

Student doesn't understand what (or when) the tutor is talking about.

Student not familiar with the physiology lingo, at least in student's opinion.

*(Note: This category is to be preferentially picked if it applies.)*

The tutor makes a statement of physiology, and the student states he/she does not understand it.

What do I do now?

The student doesn't understand something in the instructions from the tutor.

The tutor got the student confused. (e.g., tutor's mistake)

**Student doing Repair (tutor did not understand the student)**

Student thinks tutor overlooked or has forgotten something the student typed.

**Student asks non-sequitur question OR Student is completely lost**

Student asks a non-sequitur question, possibly with backward reference, showing serious misunderstanding or lack of understanding of the material.

Student declares he or she is lost. OR The student doesn't understand a *straightforward* question.

*(Note: interesting category for replanning)*

The tutor says, "Let me remind you of <something>," and the student does not confidently remember.

**Student is hedging**

"...perhaps..." OR "...??" (OR both) *(category is literal surface strings)*

<answer> <justification for answer>

Other hedges.

**Student not answering a question**

Assume *possible* initiative any time we see a long pause with no keystrokes.

The student announces reluctance to answer.

Table 1: Classification of Student Initiatives

**Student asks an explicitly case-based question**

"In one of the cases <stmt>. Is that right?" (Note: this is also hedged)

"How is <fact> relevant?"

**Student makes an explicit backward reference**

At <previous point in session> we were talking about <whatever>.

**Request for Confirmation**

For example, "So I am correct in my thinking?"

**Other initiatives**

Administrative questions

Questions specific to the structure of the experiment (e.g., rules of the "game")

Table 1: Classification of Student Initiatives (*continued*)

## Our Classification of Student Initiatives and Tutor's Responses

There are two expert tutors represented in the 28 sessions analyzed here. In this paper we discuss the student initiatives from all 28 sessions. In order to present a clearer picture of the tutor's responses, with one less degree of freedom, however, this paper only discusses the tutor's responses from the tutor who did the most sessions (16 of the 28).

Our classification of the student initiatives is in the preceding table. The following is our categorization of the tutor's responses.

- Explain or state some material in focus.
- Defer handling the initiative: perhaps modifying the tutor's model of the student.
- Do repair, stating some material, where the student did not understand the tutor.
- Request repair: the tutor doesn't understand what the student means.
- Ask student if stuck, or still stuck.
- Acknowledge the student's understanding is correct, or state it is not correct.
- Replan part or all of the remaining session.
  - \* perhaps cover material in pieces
  - \* perhaps make a big backward reference
- Give a hint, or perhaps remind student of material already covered in the session.
- Ask the student a question. (Socratic tutoring)
- State, "you are confusing X with Y." (Declare a diagnosis)
- Invite the student to review his/her thinking with the tutor.

## Discussion of Interesting Examples

One of the first things we noticed in the transcripts of the sessions is that the students may use punctuation, if at all, in a personal way, often with minimal relationship to the generally accepted conventions of English punctuation. Thus, punctuation may provide little help in recognizing the mood or clausal structure of sentences. We do not show examples of this. Repeated punctuation (e.g., "???" or "!!!") always appeared significant. The students generally capitalize conventionally. Generally, surface clues are what seem to trigger recognition of a student initiative and of its meaning. The Hedging category in the table above has some particularly clear examples of this. It appears the students consistently flag all initiatives in some fashion, so the tutor does not have to notice a departure from the current discourse focus or make similar inferences to recognize initiatives. All examples are given with the original spelling errors, punctuation, capitalization, typographical errors, and so forth.

The following example came at the end of discussing the direct physical effects, before the reflex kicks in. The abbreviations used by the tutor and student in this example are: cc=cardiac contractility, tpr=total peripheral resistance, co=cardiac output, ans=autonomic nervous system, ca=calcium [ions], and i=increase. Note that the student flags the material he wants the tutor to respond to by saying, "I'm not sure if...." Students in our sessions did this sort of thing consistently.

- tu - One last question here...
- tu - Why did you predict that cc and tpr would be unchanged?

- st - Tpr is largely a function of arteriol constriction which takes a while to adjust to co i .
- st - Cc changes in response to ans stimulation or ca build up during tachecardia.
- st - Im not sure if 120bpm is fast enough to cause that.
- tu - Probably not.

The following is another example, starting in the middle of a tutor's production. The only abbreviation is RAP=right atrial pressure.

- tu - [ . . ] what about the rate at which blood is being removed vfrom the central blood compartmanent?
- st - That rate would increase, perhaps increasing RAP???

In our sessions, the tutors appear to have a well defined picture of what they want the student to demonstrate and what the student should be tutored on if the student does not already know. Interestingly, the mere mention by the student of certain terms not introduced into the session by the tutor is enough to trigger tutoring on the parallels between those parameters and the ones the tutor is using in this session. The parallel in the following example is one of similar values: CVP and RAP are really separate measurements. The abbreviations here are co=cardiac output, RAP=right atrial pressure, and D=decrease.

- st - So, when CO I, the central venous pressure will D?
- tu - Absolutely correct.
- tu - What variable is essentially the same as central venous pressure?
- st - RAP.
- tu - Right.

Some initiatives are quite brief, their interpretation clear, and the response is fairly obvious.

- tu - OK?
- st - No

Others are complex. In the following example, SV=stroke volume. The student in this example had previously produced a 209 word response to a question, which the tutor eventually interrupted to tell the student, "you need to be more concise in your answers."

- tu - Understand?

- st - Not fully.
- st - Isn't the amount of filling equivalent to the preload?
- st - And doesn't and increased preload invoke Starling's effect?
- st - And, most importantly, what is the difference between a length/tension effect (as occurs in Starling's) and the "change in ventricular performance (SV, force,...)" which you say is not related to Starling.

As has been pointed out by research on discourse or dialogue structure, there is always some current focus, often a nested stack of subjects in focus. The preceding example establishes material in local focus, and the tutor responded by tutoring the pieces separately, then returning to the previous course of the session. This question of whether something is in or out of the current focus, seems important in recognizing the intent of student initiatives and in deciding how to handle them. The tutor whose responses are shown in this paper responded to straight questions that were off the current topic in the briefest possible fashion and then simply returned to the previous topic with no surface flagging that the topic was changing back, as if the focus had never changed. For example, an initiative as long and complex as the preceding example got the response, "Yes." On the other hand, questions about the material currently in focus generally got more elaborate treatment. For example, the following initiative took four st/tu pairs of productions to be discussed. It became a significant topic in its own right, even though this question is not part of the "standard" material to be covered in these sessions.

- st - Does RAP increase initially with increasing CO and then taper off as CO continues to I?

### Agreement Between Raters

All 28 sessions have been independently analyzed by two raters. The first analysis, which created the categorization, picked out 110 initiatives. The second analysis picked out about 210 initiatives, including 108 of the initiatives picked out in the first analysis. We have not yet had the opportunity to do a proper analysis of agreement about the categorization of these initiatives. Of the 110 initiatives picked out in the first analysis, the number per

hour-long session ranged from 0 to 11, with a standard deviation of 3.1 and mean of 3.9 per session. The most frequent categories of initiatives in the first analysis were: straight questions about material currently in focus (20 of the 110), and the category, "I am having trouble seeing/conceptualizing/grasping this" (8 of the 110). Four other common categories (each was 7 of the 110) were: straight questions about material not currently in focus, "I'm not sure if <stmt>, the student does not understand what/when the tutor is talking about, and the student is not familiar with the physiology lingo.

In one case, it appeared the student had too little grasp of the material to be able to put together a coherent initiative. That student is one of the eight who appear twice, and in the second session the same student, who had learned the material by then, generated six initiatives. This suggests an interesting line for possible future research. It seems to us that the number and depth of initiatives rises as the student's grasp of the material rises, until at some point the student knows the material thoroughly and begins to simply answer questions, with few or no initiatives.

### Future Work

We intend to focus on the context in which the initiatives occur. It is clear that how tutors respond depends on the context of the initiative. For us in CircSim-Tutor we thus need to study how to respond. We anticipate that study of the agreement on categorization between different raters could change the description of the categories. After studying agreement on categorization, we expect to have our expert human tutors categorize the initiatives we have identified. Of course, ultimately we want to incorporate what we learn about how tutors understand the students' initiatives and how the tutors respond to them into an enhanced version of the CircSim-Tutor program.

### Summary

This work attempts to categorize student initiatives encountered in tutoring a fairly small body of material in depth. It discusses the relationship between our findings and the previous work by Graesser et al. In future work we intend to focus on the context in which student initiatives occur. In applying our

result to the design of an approach for CircSim-Tutor to use in responding to student initiatives, we need to understand the way in which the human tutor decides how to respond. The availability of the tutors represented in these sessions for extended discussion should help in trying to understand this.

### Selected References

Brennan, S.E., 1991. Conversation with and through Computers. *User Modeling and User-Adapted Interaction* 1:67-86.

Fox, B., 1990. *Human Tutorial Dialogue: Final Report*. ONR Cognitive Science Program, Contractor/Grantee Meeting.

Galdes, K.K., Smith, P.J., and Smith, J.W., Jr., 1991. Factors Determining When to Interrupt and What to Say: An Empirical Study of the Case-Method Tutoring Approach. In *Proceedings of The International Conference on the Learning Sciences, 194-202*, Evanston, IL.

Graesser, A.C., Lang, K., and Horgan, D., 1988. A Taxonomy for Question Generation. *Questioning Exchange* 2:3-15.

Graesser, A.C., and Hemphill, D., 1991. Question Answering in the Context of Scientific Mechanisms. *Journal of Memory and Language* 30:186-209.

Grosz, B.J., Pollack, M.E., and Sidner, C.L., 1989. Discourse. In M. Posner, (Ed.), *Foundations of Cognitive Science*, 437-468. Cambridge, MA: MIT Press.

Rovick, A.A., and Michael, J.M., 1992. The Prediction Table: A Tool for Assessing Students' Knowledge. *Adv. Physiol. Ed.* Forthcoming.