Synthesizing Tutorial Dialogues*

Martha Evens, John Spitkovsky, Patrick Boyle
Computer Science Department
Illinois Institute of Technology
Chicago, IL 60616
csevens@minna.iit.edu

Joel Michael and **Allen Rovick**
Department of Physiology
Rush Medical College
Chicago, IL 60612

Abstract

This paper discusses problems of synthesizing tutorial discourse in an intelligent tutoring system, Circsim-Tutor, designed to help first year medical students solve problems in cardiovascular physiology involving the negative feedback system that controls blood pressure. In order to find out how human tutors handle discourse problems we have captured both face-to-face and keyboard-to-keyboard tutoring sessions in which two of the authors (JAM and AAR) tutor their own students. This paper focusses on the ways in which tutors tell students that they have made an error. We describe a classification scheme for negative acknowledgments and examine the frequency with which different types of acknowledgments occur in face-to-face and keyboard-to-keyboard sessions. Our tutors seem to make more explicit negative acknowledgments than do the tutors studied by Fox, but their acknowledgments often lead into hints that help the student continue forward in the problem-solving process. We have collected initial data about the ways in which our tutors combine hints and negative acknowledgments.

Introduction

We are building an intelligent tutoring system that carries out a tutorial dialogue with first year medical students, helping them to understand the negative feedback system that controls blood pressure, guiding them in building a qualitative, causal mental model of the system. With the goal of understanding how human tutors generate tutorial dialogues in this situation, we have captured both face-to-face and keyboard-to-keyboard tutoring sessions, each lasting an hour or more. The tutors are professors of physiology at Rush Medical College; the students are first year medical students from their classes. Analysis of the transcripts of these sessions is the basis of our attempts to generate tutorial discourse, but we find a number of serious problems as we attempt to produce natural dialogue.

The tutorial repair processes described by Fox (1988) seem extremely complex. Therefore, we are attempting to avoid repair by studying the sources of repair situations and avoiding them. In our transcripts the most common source of conversational misunderstanding is vague "how" questions from the tutor. We are trying to generate more specific questions.

Recognizing student initiatives and figuring out how to respond to them can be very difficult. At this point we can respond to requests for basic information ("What is cardiac contractility?" or "I don't understand cardiac contractility."), but more complicated initiatives are met with confusion on the part of the system (" I'm sorry, I don't understand you. Please rephrase.") Investigating tutor responses to student initiatives, we found that tutors always respond to these initiatives to some extent (Sanders et al., 1992). An initiative from the student that is relevant to the tutor's current agenda results in a modification of the plan to incorporate the issue raised by the student. Other student initiatives evoke only a brief response, followed by a return to the tutor's agenda. Revelations of serious misconceptions change the tutor's agenda to elimination of the misconception.

*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.

One of the first problems that became clear as students actually used our system is that the negative acknowledgments it delivers are crude and heavy-handed. A study of negative feedback in our human tutorial dialogues showed that our tutors use a wide variety of tactics. We describe ten categories of negative acknowledgments below and give the frequencies with which they appear in four face-to-face and four keyboard-to-keyboard sessions. Our tutors are much more likely to deliver explicit negative information to their students than those studied by Fox (1988). They use a direct negative (such as "No" or "Wrong") about 25% of the time as well as direct contradictions of what the student has just said another 10% of the time. Recent work on hints (Hume et al., 1993) has revealed that our tutors often combine negative acknowledgments with hints. We describe these combined structures briefly.

Negative Acknowledgments

The goal of this project was to look at human-to-human tutoring sessions and determine how negative acknowledgments are made. Whatever the findings, we wanted to design the natural language interface for Circsim-Tutor to reflect them as much as possible.

Methodology

We began by selecting four face-to-face and four keyboard-to-keyboard sessions that were conducted using the protocol used in Circsim-Tutor to enable consistent analysis and comparison. The face-to-face sessions were audio-taped and transcribed by a secretary. The keyboard-to-keyboard sessions were captured using our own Computer Dialogue System (Li et al., 1992). The tutor and the student were seated at PC's in different rooms; the PC's were connected via a telephone line using the Hayes Smartcom III package. The students were currently taking the course that the system is designed to support; they were fully aware that the tutor was one of the professors teaching the course. Using the two types of sessions, face-to-face and keyboard-to-keyboard, based on the same protocol enabled us to compare how tutors and students handle negative acknowledgments in the two types of sessions.

We began by identifying all negative acknowledgments in the eight sessions (F4-F7 and K25-K28). After a lot of discussion we set up the ten categories of negative acknowledgments shown below. The next step was to classify the acknowledgments using these categories (Spitkovsky & Evens, 1993).

Categories are ranked according to severity starting with error response category 1, the most severe negative acknowledgment, and ending with category 10, the least severe. In case of doubt, the classifier was instructed to choose the more negative of two adjacent categories. Some of the difficulties in classification arise because negative responses can be in response to all or part of a student's question/answer/statement. There may also be multiple error response categories contained in one "continuous" response by the tutor. This is certainly not a true one-dimensional scale, however. We are particularly interested in Category 4, which contains comments by the tutor on the language used by the student.

Negative Acknowledgment Categories

1. Direct Negative Response

The tutor responds with a direct negative remark containing an obvious negative keyword, such as *Wrong, No, Nope, Incorrect.* This category tells students they are wrong in the most severe way. The negative keyword is one of the first words in the response, often the very first. It is often followed by a question or an explanation. For example:

Tutor: Do you know a formula that gives you a deterministic statement about mean arterial pressure?

Student: Diastolic volume minus end systolic volume?

Tutor: No, that would tell you stroke volume.

2. "Indirect" Direct Negative Response

This is the same as the direct negative response, but without the negative keyword. For example:

Student: because your body always needs the same amount of blood with oxygen.

Tutor: Well, first of all, what you said isn't really correct.

3. Procedural Correction

The student is going about something wrong. Any procedural error should be placed here regardless of severity. For example:

Tutor: (Asks what parameter student will predict next.)

Student: SV

Tutor: In order to predict a parameter, you have to have to predicted its determinants.

4. Confusion Condemnation

The student seems confused and is given a distinctly negative response by the tutor. Negative remarks about a problem with language, or a lack of distinction of different terms are classified here. For example:

Student: It seems to me with sympathetic you want more blood flow.

Tutor: With sympathetic you want more blood flow? I don't know where the want comes from!

5. Repeat Student Answer - Usually Followed by a Statement or Explanation

This category carries a strong negative implication while also providing an opportunity to continue the session without the abrupt change in the course of conversation characterized by the above categories. For example:

Tutor: What does cardiac output do to the volume of blood in the central venous compartment?

Student: It's increasing it.

Tutor: It's increasing it? It seems to me that every time the heart beats, it's pulling a stroke volume of blood out of the central venous compartment.

6. Ask a Question or Make a Statement Giving a Negative Implication

This category leads students to the answer in a way that encourages student discovery. It can be difficult to tell if the acknowledgment is really negative, since an instance of this can be disguised as an answer clarification. This is the category in which to classify "repairs," where one person does not understand the other and an attempt is being made to reach an understanding.

Tutor: Well, what is the reflex attempting to do?

Student: It's attempting to lower the heart rate, I would imagine back to normal.

Tutor: Is it the heart rate that's under control?

Student: No, obviously not.

7. Repeat Current or Previous Student Answer and Ask Question Based on it

This is the first category where it may not be immediately clear to the student when he or she is wrong, since the tutor could just be testing the student's confidence in his or her answers.

Tutor: (After an extended discussion) You predicted that cardiac output would stay the same. You predicted TPR would go up. You predicted MAP would be zero. Is that possible?

8. Tutor Steers Student to Discover He or She is Wrong

The tutor doesn't immediately tell the student he or she is wrong. The tutor may just want the student to justify an answer. For example:

Tutor: What did you predict cardiac output would do?

Student: I would say it wouldn't change.

Tutor: Why do you say that?

9. Prompt for Additional Information

The student has not made a complete answer and is prompted for more, when the student is on or close to being on the right track.

Tutor: What determines the pressure in the central venous compartment?

Student: The amount of blood that's in there.

Tutor: And?

Student: I guess the peripheral resistance.

10. Minor Clarification by Tutor

In this category the answer is mostly correct or just unclear and is clarified with a statement by the tutor. The session usually continues as if the answer were correct.

Tutor: Do reflexes fully compensate for a disturbance?

Student: I don't know. That's what I was predicting, but maybe not.

Tutor: They do not. (The tutor then moved on to the next topic.)

Results

There are 103 negative acknowledgments in the face-to-face sessions and 36 in the keyboard-to keyboard sessions, for a total of 139. If you look only at the number of negative acknowledgments per hour, then you may conclude that negative acknowledgments occur more often in the face-to-face sessions. This is the wrong way to look at the data, we think. It seems more reasonable to look at the number of negative acknowledgments per turn. There are three to four times as many turns in an hour long face-to-face session as in an hour long keyboard-to- keyboard session. (These sessions behave like those we reported in Seu et al., 1991). If we look at the number of acknowledgments per turn, we see that explicit verbal acknowledgments occur more often in the keyboard sessions.

Discussion

It is not clear whether these categories are stable or not. We need to examine more transcripts to see whether all the negative acknowledgments that we find fit into these categories. Clearly, we need to perform a serious inter-rater reliability study as well. So far, another member of the team has gone through the tutoring sessions and used the above categories to classify the negative acknowledgments. The result showed 67% agreement. This leaves disagreement about 33% of the responses, almost all involving adjacent categories. We hope that this means that we can collapse some categories.

Hinting

Our tutors (JAM and AAR) feel that hints are an essential part of tutoring, and they make frequent use of this strategy. We set out to understand the hinting process better so that the Circsim-Tutor might simulate hinting. We began by analyzing the transcripts of nine two hour long keyboard sessions.

Analyzing the transcripts of nine tutoring sessions resulted in two broad hint categories. Hints either directly convey information to the student

Table 1: Negative Acknowledgments in Face-to-Face and Keyboard Sessions

	Face-toFace Sessions						Keyboard Sessions						
CAT	F4	F5	F6	F7	FTOT	FPER	K25	K26	K27	K28	KTOT	KPER	TPER
1	7	7	2	11	27	26%	2	2	0	3	7	19%	24%
2	0	2	2	3	7	7%	1	0	0	0	1	3%	6%
3	1	1	1	1	4	4%	4	2	2	0	8	22%	9%
4	1	1	0	4	6	6%	0	0	0	0	0	0%	4%
5	0	2	0	0	2	2%	0	0	0	0	0	0%	1%
6	7	12	3	4	26	25%	1	4	1	2	8	22%	24%
7	1	1	0	2	4	4%	1	1	2	2	6	16%	7%
8	8	4	2	1	15	14%	0	0	1	0	1	3%	12%
9	0	1	0	0	1	1%	0	0	0	0	0	0%	1%
10	5	3	2	1	11	11%	2	1	1	1	5	13%	12%
TOT	30	34	12	27	103		11	10	7	8	36		

(ci-hints) or point to information (pt-hints). These two hint categories may be further broken down.

CI-Hinting Categories

A1 Summary and Question

This kind of hint is a review of previous material in the tutoring session followed by an explicit question.

A2 - Summary and Implied Question

Same as A1 except that there is an implicit question in the hint.

B1 - Explanation and Question

This kind of hint is an explanation of information in the knowledg base followed by an explicit question.

B2 - Explanation and Implied Question

Same as B1 except that there is an implicit question in the hint.

C1 Explanation/Summary and Question

This kind of hint contains a summary of previous dialogue along with an explanation of information in the knowledg base followed by an explicit question.

C2 - Explanation/Summary and Implied Question

Same as B1 except that there is an implicit question in the hint.

PT-Hinting Categories

D - Question, Explicit or Implied

In this hinting style, the tutor asks a question in hopes that the student can discover the solution to the problem using known information.

E - Reply to Student Question with a Question

In this category, the tutor tries to get the student to answer his or her own question.

F Partial Acknowledgment (positive or negative)

The student has partially answered the question, but has either forgotten something or incorrectly stated part of the answer.

G - Summary of Implications of Incorrect Student Response

This kind of hint uses the student's responses to show that following this logic contradicts some previously established rule or fact.

Combining Hinting with Negative Acknowledgments

Hume and others (1993) studying hints in our transcripts became convinced that our tutors often combine negative acknowledgments and hints. To discover how these processes interact, we began by identifying the negative acknowledgments in the nine keyboard sessions used in our initial research on hints.

Table 2 shows the frequency counts for turns in which negative acknowledgments and hints co-occur. An extra column called No NA (No Negative Acknowledgment) has been added to record the number of hints that occurred without negative acknowledgments. This category includes hints that appear in response to a student question. The ColTotal row gives column totals. For all but the last column this number can be thought of as "total combined hints and negative acknowledgments." An extra row called No Hints has been added to record the number of negative acknowledgments that did not lead to a hint. This category includes those negative acknowledgments where the student was so off the mark the tutor decided an explanation was in order or the student was almost right and only a slight clarification is needed. There are also occasions when the tutor stops hinting because the student seems to find hints confusing.

Table 2: Combining Negative Acknowledgments and Hints

	Negative Acknowledgment Categories											
Hint Cat	1	2	3	4	5	6	7	8	9	10	RowTotal	No NA
A1	0	0	1	0	0	4	3	0	0	0	8	7
A2	0	0	0	0	0	0	1	0	0	0	1	0
B1	14	8	3	1	0	8	0	3	4	0	41	14
B2	2	4	2	2	0	0	0	3	0	0	13	2
C1	0	0	0	0	0	0	2	0	0	0	2	1
C2	0	0	0	0	0	0	0	0	0	0	0	3
D	3	2	5	2	0	18	0	6	8	0	44	39
E	0	0	0	0	0	0	0	0	0	0	0	2
F	0	0	0	0	0	0	0	0	10	0	10	0
G	0	0	0	0	0	3	3	0	0	0	6	1
ColTotal	19	14	11	5	0	33	9	12	22	0	125	69
No Hints	19	6	7	0	0	22	2	5	3	8	72	

Results

As Table 2 shows, there are 125 cases where hints and negative acknowledgments are combined. In the eight keyboard-to-keyboard tutoring sessions analyzed here (each two hours in length, our numbers K30–K38), there are 197 negative acknowledgments and 194 hints. Thus, if we look only at negative acknowledgments for a moment, out of the total of 197, 125 (63%) were combined with hints.

Hinting in a tutoring session can occur after a negative acknowledgment or in response to obvious student confusion or an explicit student initiative. Therefore, many hints were not associated with a negative acknowledgment. Equally, negative acknowledgments do not always lead into hints. This is because the tutor can give a negative response and follow it up with an explanation or just a simple statement of fact.

The table also makes clear, in case it was not obvious a priori, that we need to analyze much more data. The data used here is from only nine two-hour tutoring sessions so there are limited occurrences of many of the hint and negative acknowledgment categories. In fact, in the tutoring sessions used there were no occurrences of negative acknowledgment type 5 (Repeat Student Answer).

Discussion of the Relationship betwen Hints and Negative Acknowledgments

In spite of the limited range of this study there are a number of observations that can be made.

1. There is a link between Negative Acknowledgment Category 9 (Prompt for Additional Information) and Hinting Category F (Partial Acknowledgment). All occurrences of Partial Acknowledgment hints were also categorized as negative acknowledgments of type 9 (Prompt for Additional Information). Apparently, a partial acknowledgment is a negative acknowledgment asking for additional information. This analysis seems to suggest that the machine tutor should consider giving a partial acknowledgment hint whenever the student modeler detects that the answer just given is incomplete.

2. Negative Acknowledgment Category 10 (Minor Clarification by Tutor) is not used with hints.

A minor clarification by the tutor usually consists of the tutor making a brief explanation and moving to the next topic. We initially categorized several minor clarifications as being linked with a variety of hint types, but on a subsequent pass through the tutoring sessions it was determined that although the minor clarification and hint occurred in the same tutor turn, they were actually two separate responses. The minor clarification explains a minor problem and the hint is the first step in a move to the next topic of discussion.

3. Hinting type D (Question, Explicit or Implied) seems to be used after any kind of negative acknowledgment (except negative acknowledgment type 10). Once an incorrect student response has been recognized, it makes sense that no matter how severe the mistake, a question attempting to get the student to realize the mistake on his or her own can be appropriate.

4. Hinting type E (Reply to a Student Question with a Question) is not used with negative acknowledgments. This observation is certainly debatable due to the very limited amount of data analyzed so far. The argument for it can be summarized as follows: Hinting type E, by definition, is the tutor replying to a student question with a question. It does not appear that a student question would ever be answered with a negative acknowledgment, since one of the tutor's goals is to

encourage student questions. Therefore, we suspect that this observation will hold true when more data is analyzed.

5. Hinting type G (Summary of Implications of Incorrect Student Response) seems to be linked with negative acknowledgment types 6 and 7. Negative acknowledgment type 6 (ask question with negative implication) or type 7 (repeat current or previous answer and ask question based on it) both apply to summarizing or implying the student answer is wrong, which is exactly what hinting type G does. Therefore, hinting type G can only be used after a negative acknowledgment type 6 or 7. So this observation implies that hinting type G is a subset of negative acknowledgments type 6 and 7.

6. Negative Acknowledgment type 7 only uses hinting types A1 and A2 (Summary with question), C1 and C2 (Summary and Explanation with Question), or G (Summary of implications of Incorrect Student Response) when a hint is used. Negative acknowledgment type 7 by definition contains some type of summary. Therefore, only hinting types which include a summary can be used with this negative acknowledgment.

7. Negative Acknowledgment type 6 can use any hinting type, except type E (Reply to a Student Question with a Question) and type F Partial Acknowledgment).

This observation is not surprising since any hinting type which asks a question can be used after negative acknowledgment type 6 (ask question with negative implication).

Conclusion

While our tutors use a large variety of negative acknowledgment strategies, they clearly use more explicit negative acknowledgments than the tutors studied by Fox. Where could we look for an explanation of these differences? There is certainly a difference in the social situations underlying these studies. In the case of the Fox study the tutors are graduate students hired to help undergraduates through a physics course. Our tutors are professors tutoring students who are taking a course from them that covers this same material. The tutors are also the employers in our situation. The educational situations are also very different. The students in our study are older than those Fox observed; they are learning material that is essential to their performance as professionals. Our tutors are also more experienced tutors, and we conjecture that experienced tutors are more likely to give explicit negative acknowledgments.

A set of error response categories has been established along with some distribution information. Our tutors use somewhat more negative acknowledgments per turn in keyboard sessions than they do in face-to-face sessions. Our results, preliminary as they are, also suggest a strong relationship between negative acknowledgments and hints.

What does all this mean for natural language tutoring systems? We hope that we can use these results to provide a more natural interface. Clearly, the instructional planner in a tutoring system needs rules for negative acknowledgments, rules for making hints, and rules that combine these processes.

The next step in this research is to to analyze more tutoring transcripts. We have 25 more keyboard-to-keyboard transcripts full of negative acknowledgments and hints that are not yet analyzed. We have already started to try to answer the most obvious question: when do our tutors use a particular kind of negative acknowledgment? Does the severity of the negative acknowledgment correspond to the seriousness of the error? If this is so, the relationship is not immediate and obvious.

References

Fox, B. 1988. Cognitive and Interactional Aspects of Correction in Tutoring. Technical Report #88–2, University of Colorado.

Hume, G.D., Michael, J.A., Rovick, A.A., Evens, M. 1993. The Use of Hints as a Tutorial Tactic. These Proceedings.

Li, J., Seu, J., Evens, M., Michael, J.A., and Rovick, A.A. 1992. Computer Dialogue System (CDS): A System for Capturing Computer-Mediated Dialogues. *Behavior Research Methods, Instruments, and Computers (Journal of the Psychonomic Society).* 24(4): 535–540.

Sanders, G., Evens, M., Hume, G., Rovick, A.A., Michael, J.A. 1992. An Analysis of How Students Take the Initiative in Keyboard-to-Keyboard Dialogues in a Fixed Domain. In *Proceedings of the Cognitive Science Society,* 1086–1091. Los Alamitos, CA: IEEE Computer Society.

Seu, J., Chang, R-C., Li, J., Evens, M., Michael, J.A., and Rovick, A.A. 1991. Language Differences in Face-to-Face and Keyboard-to-Keyboard Tutoring Sessions. In *Proceedings of the Cognitive Science Society,* 576–580. Los Alamitos, CA: IEEE Computer Society.

Spitkovsky, J.A., and Evens, M. 1993. Negative Acknowledgments in Natural Language Tutoring Systems. In *Proceedings of the Midwest Artificial Intelligence and Cognitive Science Society,* 41–45. Carbondale, IL: SIU.