Acknowledgments in Tutorial Dialogue

Stefan Brandle and Martha Evens

Department of Computer Science & Applied Mathematics
Illinois Institute of Technology
10West 31st Street, Room 236
Chicago, IL 60616-3793
{stefan,csevens}@steve.csam.iit.edu

Abstract

In this paper we discuss an imporant aspect of human tutorial dialogue –the use of acknowledgments–and present a critique of earlier work on the use of acknowledgments in synthesizing tutorial dialogues for Intelligent Tutoring Systems (Evens et al, 1993). Our goal is to establish a more solid theoretical base for studying and synthesizing tutorial dialogues than what was used in the earlier work. A proposed foundation for this study is the idea of language use as a joint activity composed of joint actions, presented by Clark (1996). It appears to provide a more powerful conceptual linguistic framework within which to describe the behavior of human tutors and shows promise as a guide for synthesizing tutorial dialogues in Intelligent Tutoring Systems. In particular, we are investigating the role of acknowledgments as the mechanism that makes joint action possible. It is our thesis that an acknowledgment is anything that signals closure or lack of closure of a joint action.

Introduction

We have built an intelligent tutoring system, CIRCSIM-Tutor, that carries out a natural language (text-based) tutorial dialogue with medical students. The system goal is to help the students develop a better mental model of the mechanisms that control blood pressure in human physiology. The system design is based on the analysis of transcripts of medical professors tutoring their students in this domain. We have a running system that models the behavior of our human tutors and are now studying the transcripts in order to develop a better understanding of all aspects of successful tutoring. The goal is to use the linguistic and tutoring knowledge gleaned from the transcripts in order to improve the linguistic and pedagogical quality of CIRCSIM-Tutor. This paper deals specifically with acknowledgments, an important low level mechanism in text generation, text understanding, lesson planning, and student modeling.

In order to develop a better understanding of acknowledgments, we need a definition and theory that will prove productive in the enhancement of our tutoring system. Previous work by Evens et al (1993) proves useful, but is too ad hoc to support long-term continued enhancement. Papers by Fox (1988) and Graesser (1993) discuss acknowledgments while analyzing repair and feedback, but do not explicitly develop a theory of acknowledgment.

We are proposing that an acknowledgment is anything that signals closure or lack of closure of a joint action. The surface form of the signal can vary widely, but there are standard features of acknowledgments that make it possible to study them and enhance the ability of an intelligent tutoring system to use them.

Theoretical Framework

We have chosen to define acknowledgments in the light of Clark's presentation of discourse as a joint activity composed of joint actions (*Using Language*, 1996). He focuses on the mutual coordination of their individual actions by the participants in linguistic activity. Clark says: "There is coordination of both *content*, what the participants intend to do, and *processes*, the physical and mental systems they recruit in carrying out those intentions" (p. 59). Similar ideas are expressed by Fox (1993) who states as a major finding of her book, *The Human Tutorial Dialog Project*, that "tutoring involves constant, and local, management. This requires a pervasive mutual orientation between tutor and student, such that every session (indeed, every utterance) is a thoroughly interactional achievement, produced by both tutor and student" (p. 3).

Clark says that joint actions can be divided into phases entry, body, and exit and the phases are what get coordinated. Entries and exits are coordinated by syntactic, morphological, and intonational markers. A joint action is complete when there is a mutual recognition of closure on that action. "It is a fundamental principle of intentional action that people look for evidence that they have done what they intended to do." (p. 222) He restates Norman's *Principle of closure*: "Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it." (p. 222) and then introduces the *Principle of joint closure*: "The participants in a joint action try to establish the mutual belief that they have succeeded well enough for current purposes" (p. 226).

This model states that there are concurrent joint actions taking place on four separate levels. From the lowest level

to the highest, they are: 1) Speaker presents a signal and addressees attend to signal. 2) Speaker signals something to addressees, who identify the signal. 3) Speaker signals that *p*, and addressees recognize that the speaker means *p*. 4) Speaker proposes a joint project and the addressees take up the proposal. (Clark, pp. 148-151). The action at each of these levels is performed in order to make possible the action at the next higher level, i.e. the speaker presents a signal in order to signal something, in order to signal that something, in order to propose something. The addressee attends to the signal in order to identify the signal, in order to determine what the speaker means to communicate, in order to be able to take up the proposal.

Clark (and others, e.g. Stubbs, 1993) divide communication into two tracks or channels. The first track is the one in which the business is transacted, such as the tutor requesting that the tutee make a prediction about the value of cardiac output and the student responding. The second track is the metacommunication track is the one in which the participants monitor the success of the communication, make adjustments in the communication parameters, detect and signal communication problems, and attempt to fix communication problems. Stubbs states that metacommunication is for "attracting or showing attention, controlling the amount of speech, checking or confirming understanding, summarizing, defining, editing, correcting, specifying topic" (pp. 50-53).

One of the categorizing features of an acknowledgment is the degree to which it is explicitly an acknowledgment vs. playing that role implicitly. If the utterance is "Good, your answer is correct.", that is explicitly a positive acknowledgment. However, the utterance could instead be "Next, let's talk about cardiac output." In this case, the implication is that the answer was correct, but that is never explicitly stated and must instead be derived from the context and the lack of a negative acknowledgment. A related feature is the presence or absence of a keyword that functions as a discourse marker to indicate an acknowledgment. Examples are words like "yes" and "wrong".

It is worth mentioning that this proposed framework fits very closely with the acknowledgment protocols that are used effectively in artificial communication protocols (Brandle, 1996). This suggests that acknowledgment devices are central to communication in general.

Application to Tutoring

In our keyboard-to-keyboard tutoring environment, the layers are as follows: 1) The tutor presses the keys on his keyboard, causing material to be displayed on the tutee's computer terminal and the tutee is attending by looking at the screen. 2) The tutor is typing a sentence to the tutee who reads the characters, forming words and putting them together into an utterance. 3) The tutor is presenting the request "State what happens to cardiac output when heart rate increases" and the tutee understands that the speaker wishes him to determine and utter an answer. 4) The tutor is requesting that the tutee decide what happens to cardiac output when heart rate increases and inform the tutor of the conclusion; the tutee then takes up the joint project.

Each joint action in the metacommunication track is itself a regular joint action with all the characteristics of joint actions; it is a joint action whose function is to address and redress issues arising out of other joint actions. Stubbs states that "metacommunication highly characteristic of teacher-talk, not only because it comprises a high percentage of what teachers do spend their time saying to their pupils, but also in the sense that its use is radically asymmetrical. Speakers hold quite specific expectations that it is the teacher who uses it. It is almost never used by the pupils..." (p. 53) An example is the following level one timeout (the tutor [*tu*] saw nothing from the student [*st*] and performed a tutor interrupt during the tutee's keyboard turn marked as [*ti*] in the transcripts to determine what was happening):

K30-tu-2-3: Please read page 1 in the notebook.
K30-st-3-1: { TU: DO YOU NEED HELP?}
K30-st-3-2: {INTERRUPTED STUDENT INPUT}
K30-ti-4-1: Please enter your social security number and then press <enter>
K30-st-5-1: XXX-XX-XXXX
K30-tu-6-1: Remember that you need to press enter to end each of your turns.

Another example of metacommunication occurs in the following transcript fragment. The tutee doesn't understand the term "CC" and initiates a repair side sequence to solve the problem.

K30-tu-78-2: Now take a look at your predictions.
K30-tu-78-3: You said that RAP would increase and CC would decrease.
K30-tu-78-4: What will happen to SV?
K30-st-79-1: Stroke volume increases, but what's CC
K30-tu-80-1: Sorry I misspoke in my last message.
K30-tu-80-2: You said that RAP would DECREASE and C(ardiac) C(ontractility) would INCREASE.

Acknowledgments in Tutoring

1) Positive vs. Negative:

Positive acknowledgments mark closure on a joint action.

K25-tu-50-1: And what determines cvp?
K25-st-51-1: Blood volume and "compliance" of the venous side of the circ.
K25-tu-52-1: Right.

Negative acknowledgments signal a failure to reach closure on a joint action.

K27-tu-28-1: Next [prediction]?
K27-st-29-1: Sv
K27-tu-30-1: In order to predict sv you need to have predicted its determinants.
K27-tu-30-2: Try again.

2) Explicit vs. Implicit

Explicit acknowledgments are such that a surface-level inspection clearly signals the positive/negative acknowledgment value of the utterance. In the case of implicit acknowledgments, the acknowledgment value must be derived or calculated from the context.

K25-tu-52-1 above is an instance of an explicit positive acknowledgment with a marking keyword. The tutor clearly marks the tutee's answer as correct and signals closure on the joint action of establishing that the tutee knows what determines cvp in that context.

K27-tu-30-1 is an implicit negative acknowledgment since it doesn't directly state that the answer "Sv" is not acceptable. It requires that the tutee build a syllogism (predict the determinants of a variable as a necessary condition for predicting the variable), recognize the implicit denial of the consequent, and conclude that the prediction was not being accepted and why that is the case.

3) Communication Level Involved

There are multiple joint projects active in parallel at the different levels. Acknowledgments are needed to mark closure or lack of closure for each of these joint projects. A relevant concept is the property of downward evidence (p.148), *evidence that one level is complete is also evidence that all levels below it are complete*. In other words, a positive acknowledgment at level three understanding implies a positive acknowledgment at layer two (the tutee recognized the signal) and a positive acknowledgment at layer one (the tutee was paying attention). This makes sense in view of the principle of least effort, "All things being equal, agents try to minimize their effort in doing what they intend to do." (Clark, p. 224) It is generally not possible to succeed at the level of uptake on the joint action (level four), without having understood the signal (level three), having decoded the signal (level two) and paid attention to the signal (level one). Consequently, a positive acknowledgment at one level also functions as an implicit acknowledgment at all lower levels. It is worth noting that this implies that positive acknowledgments should be issued at as high a level as possible to obtain maximum closure on lower levels and thus maximize efficiency. It also implies that negative acknowledgments are issued at the lowest possible level. In normal situations (e.g., not performing face saving), if the listener were unable to perform an uptake on the joint action (level four) because of a failure to decode the signal (level two), it would not make sense to say "I disagree" or "I refuse" (level four) and lead the speaker to focus on level four when the problem is really at level two. In the same situation, it could also be inefficient to signal a level one problem ("I'm sorry, what were you saying?") if the problem was one that could not be solved by repeating the signal (e.g., the listener cannot decode Spanish). It is worth noting that when the speaker does not otherwise give any clues about the acknowledgment value of an utterance, it generally defaults to being an implicit positive acknowledgment.

4) Track One vs. Track Two

In the example K30-st-79-1: "Stroke volume increases, but what's CC" the tutee performs a track 1, level four uptake ("Stroke volume increases") and a track two, level three negative acknowledgment ("but what's CC"). The problem is the on-the-fly abbreviation of "cardiac contractility" to "CC". The tutee had a problem understanding the language used. Other track 2 examples are things such as the tutor reminding the tutee to press the return key when done typing.

Application to Prior Work

Evens et al. (1993) presented a study of negative acknowledgments and hinting in "Synthesizing Tutorial Dialogues." Their motivation stemmed from Susan Chipman's observation (reported in Spitkovsky & Evens, 1993) that Fox's tutors essentially never said "Wrong" or "No", whereas our system handed them out all the time. Evens et al. showed that the expert tutors used explicit negative keywords, but only about 25% of the time; 75% of the negative acknowledgments were of a different form. They developed an ad hoc set of ten categories of negative acknowledgment which were ranked according to severity, from most severe (direct negative response) to least severe (minor clarification by tutor).

The study had a problem in that there was no serious reference theory of acknowledgments to guide the development and ranking of the categories, or even to support deciding what utterances were negative acknowledgments. This lack of a strong theory also posed a problem for attempts to rank negative acknowledgment categories on a severity scale. The authors recognized weaknesses and stated "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" (p. 138). It is our belief that for many of those negative acknowledgments the differences are more qualitative than quantitative, and that the attempt to rank them on a severity scale is misguided. Granted, it is true that some negative acknowledgments are perceived as more severe than others (e.g., K25-tu-114-1: "No, I think you are reasoning backwards" is more severe than K25-st-62-1/2: "So, in dr hr is up, co is up, but sv is down. How is this possible?"), but this doesn't demonstrate that categories of negative acknowledgment are more severe than others. In general, the perceived "negativity" of a negative acknowledgment depends more on the context and the role the acknowledgment plays than on the surface form of the

acknowledgment.

The features that we propose as providing a basis for characterizing acknowledgments are the following: 1) positive vs. negative, 2) explicit vs. implicit, 3) track one (normal communication) vs. track two (metacommunication), and 4) the level at which it is used (levels one through four) and 5) the presence of a keyword that functions as a linguistic marker for acknowledgments. These are the values that a text-generation system would have to consider in choosing what type of acknowledgment to issue.

An analysis of the previously-proposed ten categories suggests that they describe common techniques for issuing negative acknowledgments, but that the categorization is 1) by no means orthogonal and 2) provides limited usefulness to a text-generation system which must decide what category of acknowledgment to issue. One of them, Direct Negative Response, is categorized strictly on the basis of whether an explicitly negative keyword is present in the utterance. Another, "Indirect" Direct Negative Response, covers all explicit negative utterances in track one or two, levels one through four, which do not contain a negative marking keyword. A more useful categorization from the perspective of text generation would provide better guidance in determining what utterance to issue. For instance, the system knows that the tutee made a mistake in predicting that blood pressure would rise; it would not change. This indicates a negative acknowledgment at level 4 in track 1, but there remain the questions of whether this should be explicit or implicit, have an explicitly negative keyword, and deciding which of the categories meeting the (negative acknowledgment, level 4, track 1) criterion n-tuple should be used. We can not, unfortunately, just match features to choose a category of acknowledgment, because some of the ten categories differ on the basis of pedagogical purpose and technique, not just syntactic and semantic features.

Implicit negative acknowledgments may be better pedagogically in that they force the tutee to work at a higher cognitive level in order to decode the tutor's utterances. Another case where implicit acknowledgments are preferred is the matter of what counts as sufficient evidence from the student for the tutor to conclude that the student knows the material. Our expert medical tutors avoid asking yes/no questions such as "Do you understand?" on the basis that the self-evaluation returned by the students is often faulty. This position is reinforced by the findings in (Graesser, 1993) that "the most reliable information source for inferring student knowledge was the students' answers to questions" (p. 128). Our tutors would rather ask questions to which the student cannot get away with issuing a positive acknowledgment, but must instead furnish evidence which can be construed as sufficient proof of understanding to achieve closure on the joint action at hand. Clark presents four classes of signals that are positive evidence of understanding: 1) assertions of understanding, 2) presuppositions of understanding, 3) displays of understanding, and 4) exemplifications of understanding (pp. 228-229). He states that "displays and exemplifications tend to be more valid evidence than assertions and presuppositions."

Application to Intelligent Tutoring Systems

It turns out that the idea of language use as a joint activity composed of joint actions maps very nicely onto the sort of planning and linguistic activity that must be carried out by a natural language based ITS. In particular, this provides a very nice framework for deciding when to issue acknowledgments and guidance on what the text generation processor should issue. In the case of CIRCSIM-Tutor, which very closely controls the dialogue, this turns into the system presentating a joint project and looks for the student's uptake of the project–or refusal to do so, The system analyzes the response, and if it is sufficient to declare the goal reached, the system can issue a positive acknowledgment and continue. If the answer was wrong or insufficient, the system can issue a negative acknowledgment, or a partial positive acknowledgment. If what the student produced does not appear to be an uptake of the project, then the input can be treated as track two communication or as a student initiative and handled accordingly. We are currently working on mapping this acknowledgment rules into tutoring and discourse generation rules in the CIRCSIM-Tutor planner and plan to report on and demonstrate the resulting system in the future.

Conclusions

We believe that approaching acknowledgments in communication from a joint actions framework is a productive one that is an advance over previous efforts relating to acknowledgments. We are using this to perform a qualitative study of acknowledgments in our transcripts, expect results in the near future, and are comfortable that those results will support the proposed framework.

Acknowledgements

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.

References

Brandle, S. 1996. Understanding Human Communication Acknowledgement Protocols by Studying Artificial Communication Protocols. In *Online Proceedings of Annual MAICS Conference.*

http://www.cs.indiana.edu/event/maics96/Proceedings/brandle.html.

Clark, H. H. 1996. *Using language*. Cambridge, Great Britain:Cambridge University Press.

Evens, M.; Spitkovsky, J.; Boyle, P.; Michael, J.; and Rovick, A. 1993. Synthesizing Tutorial Dialogues. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* , 137-142. Hillsdale, NJ:Lawrence Erlbaum Associates.

Fox, B. 1988. Repair as a Factor in Interface Design. Technical Report 88-4. Boulder, Colorado: Institute of Cognitive Science, University of Colorado, Boulder, Colorado.

Fox, B. 1993. *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems.* Hillsdale, New Jersey:Lawrence Erlbaum Associates.

Graesser, A. C. 1993. Dialogue Patterns and Feedback Mechanisms during Naturalistic Tutoring. In Proceedings of the Annual Conference of the Cognitive Science Society, 126-130. Hillsdale, NJ: Lawrence Erlbaum Associates.

Spitkovsky, J. A.; and Evens, M. W. 1993. Negative Acknowledgements in Natural Language Tutoring Systems. In *Proceedings of the Annual MAICS Conference*, 41-45.

Stubbs, M. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language.* Chicago, Illinois: University of Chicago Press.