In Daniel C. Edelson and Eric A. Domeshek, eds., *International Conference on the Learning Sciences* (Proceedings of ICLS '96, Evanston, IL), 1996, pp. 75-82.

Using Tutoring Patterns to Generate More Cohesive Text in an Intelligent Tutoring System

Reva Freedman
Department of EECS
Northwestern University
freedman@delta.eecs.nwu.edu

Abstract: If we want students to read, pay attention to and understand the text generated by an ITS, the text needs to (a) be organized into cohesive turns, (b) flow in natural conversation patterns, and (c) have sufficient variety that neither syntactic structures nor pedagogical goals are repeated overly often. CIRCSIM-Tutor v. 3 is a natural-language based ITS for cardiac physiology currently under development. Text planning for v. 3 will be accomplished by TIPS, a text planning engine based on current research in text generation. We describe how TIPS will achieve the goals mentioned above. We use transcripts of sessions conducted by expert human tutors to derive tutoring patterns for TIPS. Through the use of the TIPS planner, we expect CIRCSIM-Tutor v. 3 to generate longer, more complex and more natural text than previous work.

1. Introduction

From the inception of the genre [Carbonell 1970], authors of text-based ITSs have faced a difficult choice. Since the ability to generate text in a principled manner has not developed as rapidly as the set of pedagogical ideas one might wish to express, natural-sounding text has had to be largely hand-crafted for each situation the mechanized tutor might face. But hand-crafting text, or using surface-structure templates to hard-code grammatical forms, reduces the economy of scale which an intelligent system should provide. On the other hand, the quality of the language directly affects students' ability to understand and learn from the generated text, so reducing the quality of the generated text is not a desirable option either. The goal of our research is to demonstrate how current research in text generation can be used to improve the output of a text-based ITS.

We begin by looking at the text generated by CIRCSIM-Tutor v. 2, an ITS developed by our project in 1992. A critical look at the generated text reveals several problems [1]. A major limitation of v. 2 of CIRCSIM-Tutor is the fact that it has only one way of teaching any given concept. As a result it is easy for students to memorize a concept before they truly understand it because they have simply heard it too many times. Another major limitation of v. 2 is that it does not pay attention to intra-turn or inter-turn cohesion, resulting in choppy and stilted text. Finally, v. 2 does not have the capability to change from one tutoring plan to another.

As a result of studying these problems, we propose an improved architecture for dialogue-based ITSs. We sketch the design of TIPS, the text planner for CIRCSIM-Tutor v. 3, paying particular attention to the aspects which are intended to help correct the above deficits. To identify tutoring patterns for TIPS to generate, we examine human-to-human tutoring transcripts derived from the work of expert tutors.

We expect that the combination of a new planner and a larger tutorial knowledge base will permit CIRCSIM-Tutor to generate longer, more complex and more natural-sounding text than previous work.

[1] The problems described are not specific to CIRCSIM-Tutor.

2. Important Issues in Text Generation for ITSs

2.1. Motivation for a New System

CIRCSIM-Tutor v. 2 is a text-based ITS which tutors students on the negative feedback loop which maintains a steady blood pressure in the human body, one of the more difficult topics in the first year medical curriculum. In class lectures, students are given a simplified qualitative model of the heart. When students sit down with CIRCSIM-Tutor, they are given problems to solve. In each problem, something happens to change the processing of the heart. The student is then asked to predict the direction of change of seven core variables during three stages of processing. After each phase, the tutor conducts a dialogue with the student to correct the value of any variable which had been missed.

The key to solving the problems is knowing how to propagate changes from one variable to the next. Three types of knowledge are required: the determinants of each variable, which determinant is relevant in any particular case, and how each determinant affects the value of the determined variable.

The following text, whose goal is to teach the student the correct value of the variable SV, is characteristic of v. 2. It shows a pattern which is used to correct any variable whose value is not mediated by the nervous system. [2]

1 T: What are the determinants of SV?
2 S: MAP.
3 T: MAP is the wrong answer.
SV is determined by RAP and CC.
What is the important determinant in this case?
4 S: RAP.
5 T: Correct, the important determinant of SV in this case is RAP.
What is the relation between RAP and SV?
6 S: Direct.
7 T: Good, you got the correct answer.
What is the correct value of SV?
8 S: Decrease.
9 T: Correct, the value of SV is decreased.

The tutor's pedagogical plan could be described as follows:

Correct-variable:

Get determinants of variable	*(turn 1)*
Get major determinant	*(turn 3)*
Get relationship between variable and determinant	*(turn 5)*
Get value of variable	*(turn 7)*

The generated text is grammatically accurate and pedagogically reasonable. However, the text has the stylized, stilted quality we associate with compiler error messages rather than with language used in natural situations. In the following sections, we will attempt to identify the problems in this text with greater precision.

2.2. Language-Related Limitations: Intra-Turn Cohesion

The examples below show some more natural replies which the tutor could give in turn 3. Each of them

[2] The seven core variables are: SV = stroke volume, MAP = mean arterial pressure, RAP = right arterial pressure, CC = cardiac contractility, CO = cardiac output, HR = heart rate, TPR = total peripheral resistance.

expresses the same ideas as the original, but in shorter and more natural-sounding language.

(1) T: MAP is wrong because SV is determined by RAP and CC.
Which is the important one in this case?

(2) T: No, the determinants of SV are RAP and CC.
Which one is relevant here?

(3) T: No, the right answer is RAP and CC.
Which one is most important in this case?

Studying the proposed replacements allows us to clarify some of the problems with the original text. First, v. 2 always implements the same pedagogical goal in the same way. The v. 2 planner knows one way to say each of the four goals in the plan above. While it can change a few words, such as alternating between "correct" and "good", the sentence structure and the main verb never vary. Furthermore, the choice of lexical items is made on a random basis rather than as human speakers would do it.

A more important problem is that although each sentence that the tutor utters is accurate and grammatically correct, the turns are not put together well. Each sentence uttered by the tutor expresses a proposition with all arguments fully spelled out. In normal human discourse, people leave some parts of most propositions implicit. The fully spelled-out propositions accentuate the related problem of sudden shifts in topic, such as from MAP to SV in turn 3. These issues could be summarized as a lack of *intra-turn cohesion* [Halliday & Hasan 1976]. In the suggested replacements, several linguistic mechanisms are used to increase cohesion: sentences have been combined, pronouns have been inserted where appropriate, and some ideas have been left implicit.

The issues of intra-turn cohesion and syntactic variety are closely associated. Although it would be easy to add more lexical items and sentence structures, that would just create more cases which need to be tweaked in order to produce natural-sounding text. The underlying issue is that v. 2 generates text for each pedagogical goal independently. We need a way to take into account that several goals could be uttered together in one turn without having to provide a template for every combination of goals.

2.3. Language-Related Limitations: Inter-Turn Cohesion

A related problem is that v. 2 of CIRCSIM-Tutor plans text independently for each turn without taking previous turns into account. For example, turn 7 could be expressed more smoothly as follows:

(4) T: Correct. So what is the value of SV?

The discourse marker "so" makes this text sound more natural because it reminds the reader that a subtopic has been completed and that the tutor is returning to the previous topic [Schiffrin 1987]. Because v. 2 generates text for each turn independently of the evolving conversation, it never generates such cues for the reader.

2.4. Lack of Pedagogical Variety

A final problem with v. 2 is that the pedagogical plan identified above is the only one known to the tutor. This limitation makes it too easy for the student to "go through the motions" without thinking about what the tutor is saying. Furthermore, the work done by our project on creating a more sophisticated student model is less useful without a selection of alternatives for the planner to choose from. The issue of pedagogical variety is tightly coupled with the ability to backtrack. In v. 2, if a student cannot follow the plan, the tutor has no choice but to continue trying or to give the student the answer. The provision of pedagogical alternatives and a student model which can distinguish among them would be more useful if the tutor could abandon unsuccessful lines of argument.

3. Expanding Pedagogical and Syntactic Variety

Over the past four years, the CIRCSIM-Tutor project has collected over 5000 turns of written, human-to-human tutoring sessions using expert tutors. We have analyzed these transcripts in order to identify pedagogical and linguistic strategies which v. 3 can implement. In this section, we illustrate several categories of pedagogical and linguistic patterns, not available in v. 2, which can be used to improve the quality of the generated text in v. 3.

3.1. Using More Complex Tutorial Patterns

Version 2 corrects the student's incorrect predictions one variable at a time. In addition to looking at the student's predictions, expert tutors also look at the relationships between them. For example, expert tutors may use a correct prediction as leverage to disabuse the student of an incorrect belief. We have identified two ways to carry out this operation.

In turn 3 of the example above, the tutor first informs the student that a previous answer was wrong, then provides the correct answer. One alternative frequently used by the expert human tutors in the transcripts is to ask a *pseudo-diagnostic question* such as the following:

(5) T: Can you explain the mechanism by which MAP determines SV?

(cf. K14:33 [3])

We have named this pattern the "pseudo-diagnostic question" because it has the same form as a true diagnostic question but is not intended to elicit an explanation, as there is no such relationship in this case.

The pseudo-diagnostic question is intended to disabuse the student of an incorrect idea by asking a question which the student cannot answer. A more direct approach is to use the *show-contradiction* pattern, where the tutor identifies a set of contradictory statements which the student has made. We let the student determine which prediction to withdraw. Here is an example:

(7) T: So RAP and CC determine SV.
You predicted that CC would be unchanged and that RAP increased.
How can SV be unchanged?

(K27:68)

A second way in which expert tutors make use of the relationship between predictions is in the generation of context-dependent patterns. For example, repeating or referring to the student's previous statements can provide a shorter and more natural way to move from one variable to the next, as opposed to repeating the entire *correct-variable* schema shown above as v. 2 would do. The following examples give an idea of the potential variety.

(7) T: If HR increases, which variable will be immediately and directly affected by this change?

(cf. K48:62)

(9) T: Well, you predicted that RAP would in fact go down and you predicted that CC would not change.
So what happens to SV?

(K14:53)

[3] The references indicate transcript number and turn number in our corpus.

(9) T: Well, you made predictions about how RAP and CC would change as a result of the pacemaker malfunction.
What do you think will happen to SV?

(K14:51)

3.2. Using Deeper Domain Knowledge

Human tutors frequently refer to a deeper level of domain knowledge to help the student identify determinants. Often the additional knowledge refers to a functional model of the domain, such as in the following attempt to elicit the determinants of SV.

(10) T: What I was asking is what determines how much blood is ejected from the heart each time it beats (i.e. the SV).

(K14:49)

Appeals to a functional model range from one-sentence substitutions like the one above to long explanations. The following excerpt shows an intermediate-length example. In this example, the tutor decides to move to a functional model after failing to elicit the determinants of RAP. This excerpt is part of a larger structure for teaching the relationship of RAP and CO.

(11) S: *(gives wrong determinants for RAP)*
T: Try to look at it this way. The right atrium is like the end of the venous system. And the veins are very compliant.
…
T: The pressure in any compliant structure is determined by the volume contained by that structure and the compliance of that structure.
S: Then it's CO.
T: Super!…

(K30:65–70)

3.3. Combining Patterns

Human tutors combine tutorial patterns in several ways. First, patterns can be nested. Second, when a pattern fails due to an incorrect response on the part of the student, the tutor can switch to a new pattern. Finally, error correction patterns can be used to fix up partially correct answers so that the original pattern can continue. [Fig. 1] shows examples of each of these cases as they will be implemented in v. 3.

The following pattern is the most common pattern used to correct variables controlled by the nervous system.

Correct-neural (V):
(1) Make sure student knows that V is neural
(2) Make sure that student knows that current stage is pre-neural
(3) Make sure student knows that correct value of V is no-change

[Fig. 1] shows in condensed form a set of dialogues which CIRCSIM-Tutor can generate from this schema for the variable TPR, varying only the implementation of the first subgoal. Each item in italics represents text which could be generated by one semantic form. Where convenient, we have shown two possible realizations separated by a slash. The numbers correspond to the subgoals in the schema above. The student's responses are shown in roman type. In the leftmost path, the student gives the desired answer immediately. In the second path, the student gives a partially correct answer, in this case an answer which is true but does not use the tutor's desired language. The tutor adds a goal to correct the student's language before continuing with the schema. In the third branch, the student gives an answer which is on the path toward the correct answer. The

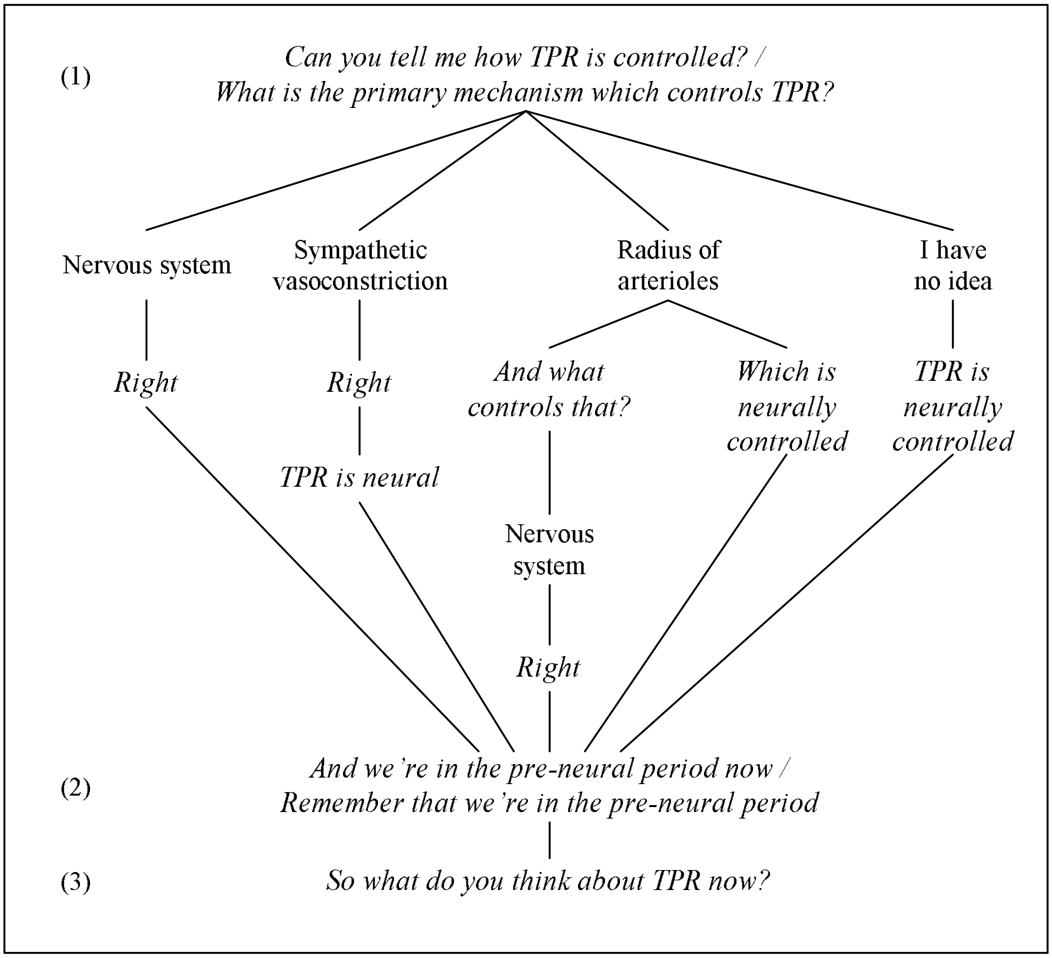

Figure 1: Conversations which CIRCSIM-Tutor can generate

tutor helps the student toward the correct answer in two different ways. In the final branch, the tutor gives the student the answer in order to avoid a long dialogue.

By varying the other subgoals of this schema, as well as utilizing the option to switch to a different schema, we can generate a large and varied set of dialogues.

3.4. Combining Concepts Inside a Turn

Each of the following turns expresses the three concepts which constitute the *correct-neural* schema. However, in (12) the turn planner has combined the first two forms into one sentence, while in (13) the last two forms have been combined.

(12) T: TPR is controlled by the nervous system, and we're talking about what happens before there are any neural changes. So TPR doesn't change.

(13) T: TPR is a neurally controlled variable…Then what value would you assign to TPR in the pre-neural stage?

3.5. Realizing a Concept in Multiple Ways

At the lowest level of detail, a semantic form emitted by the tutorial planner can be realized with multiple choices of verb and argument structure. Although each example may be simple, the availability of multiple realizations for several semantic forms in a turn can create markedly different text.

(14) T: … And what happened to RAP?

(15) T: … What effect would this have on RAP?

(16) T: … what must happen to RAP?

(17) T: … How will that affect RAP?

4. Implications for the Planner

Note that example (5) only covers the first goal of turn 3, i.e. informing the student about the error. How the remaining goals are achieved depends on the student's response to (5). While the use of additional tutoring patterns creates increased opportunities for student interaction, it means that we can no longer use a simple top-down planner to generate the text. Depending on the student's response, we might want to continue with our current plan or drop it and try another one. In the latter case, we need to be able to remove the remaining subgoals of a pattern from the agenda without losing track of turns already uttered in service of the already-satisfied subgoals. For example, suppose we wanted to terminate the *correct-variable* plan after the second subgoal. As we can't "unsay" a turn, the already-generated text, turns 1 through 4, would still be part of the conversation and would still be relevant when attempting to increase the cohesion of later turns.

For this reason, the TIPS planner uses separate modules running in parallel to keep track of pedagogical goals and the evolving conversation. The *tutorial planner* is responsible for choosing pedagogical goals and maintaining an agenda for the tutor. It receives interpreted input from CIRCSIM-Tutor's input understander, identifies the next appropriate pedagogical goal and produces semantic forms for the turn planner. Its responsibilities include making sure that each repeated attempt to respond to a student error is implemented differently from the previous attempt, whether by using a different tutoring pattern or via a different syntactic or lexical structure. Much of the work of ensuring coherence is handled automatically in this module by putting appropriate constraints on the plan operators.

The *turn planner* is responsible for combining semantic forms into turns and ensuring intra-turn and inter-turn cohesion. It accumulates semantic forms until it receives one (e.g. *elicit*) which requires a response from the student. At that point it combines the semantic forms into a cohesive turn, attaches that turn to the discourse tree, and issues the text to the student.

An overview of CIRCSIM-Tutor v. 3 is shown in [Fig. 2].

5. Conclusions

This paper describes TIPS, a new planner which will be used to generate natural-language dialogue for v. 3 of CIRCSIM-Tutor, a text-based ITS for cardiac physiology. The name stands for "Text generation Interactively, a Planning System", an acronym chosen because of the pun involved: after all, a primary goal of CIRCSIM-Tutor is to generate verbal hints to help the student, i.e. tips. A primary goal in the design of TIPS is to use current knowledge about text planning to improve the variety and cohesiveness of the generated text. Using the TIPS planner and a knowledge base derived from the study of naturalistic transcripts, we expect that CIRCSIM-Tutor v. 3 will generate text which is more natural, varied and cohesive than previous ITSs.

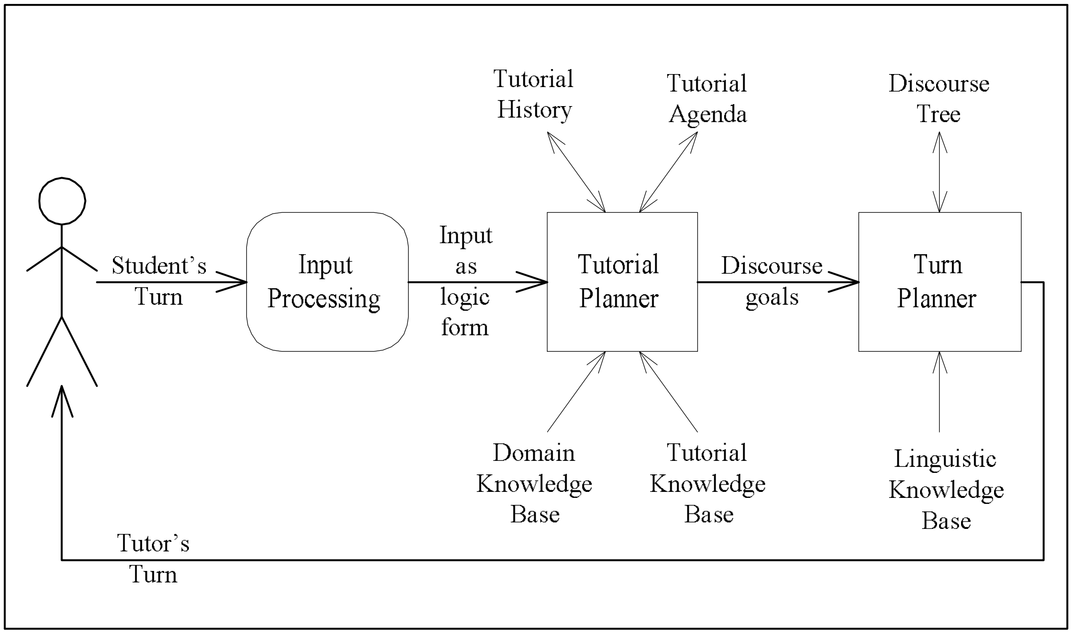

Figure 2: Overview of CIRCSIM-Tutor v. 3

References

[Carbonell 1970] Carbonell, J. R. (1970). AI in CAI: Artificial intelligence approach to computer assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202.

[Halliday & Hasan 1976] Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

[Schiffrin 1987] Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Acknowledgments

This research was conducted at the Illinois Institute of Technology as part of the CIRCSIM-Tutor project. It 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.

This work could not have been completed without the enthusiasm and practical assistance of the project director, Professor Martha W. Evens of the Illinois Institute of Technology; my advisor, Professor Gilbert K. Krulee of Northwestern University; and co-investigators and domain experts Professors Allen A. Rovick and Joel A. Michael of Rush Medical College. I would also like to thank the many members of the CIRCSIM-Tutor project team, past and present, who shared with me their ideas on how and what to tutor.