SOME PHENOMENA HANDLED BY THE CIRCSIM-TUTOR VERSION 3 INPUT UNDERSTANDER

Michael Glass
Illinois Institute of Technology
glass@charlie.iit.edu

Abstract

The CIRCSIM-Tutor version 3 input understander is the module which will process student answers during a tutoring dialogue. We have transcripts of human tutors conducting such dialogues, as well as logs of students using earlier versions of the computer tutor. This paper addresses how the input understander handles some of the language phenomena we have observed. Even when the student is giving short answers to simple questions, the tutor must sometimes handle them as simple student initiatives.

INTRODUCTION

CIRCSIM-Tutor is an intelligent tutoring system intended to tutor medical students about a negative feedback loop for blood pressure regulation in the human body.

CIRCSIM-Tutor is language-based. It instructs by carrying out a dialogue with the student in English. There are no diagrams, hypertext links, or pull-down menus. The student is first presented with a description of an event which disturbs blood pressure (e.g., a hemorrhage). The student then predicts the qualitative changes in certain physiological variables related to blood circulation (e.g., volume of blood pumped per minute). Except for a chart the student fills in to record predictions, the student types black letters on a white screen and the tutor replies in kind.

This paper discusses some of the language and implementation issues for processing student input in the next version (version 3) of CIRCSIM-Tutor [Freedman, 1996], building on the previous version (version 2) [Woo, 1991].

We have fairly extensive data on student input, both from logs of students using version 2 and from approximately 80 one- and two-hour keyboard-to-keyboard tutoring sessions. The examples in this paper have been taken from these sources. Unless necessary to make a point, they have had their spelling and punctuation corrected. An identifier such as "K40-63" locates an extract in our corpus. "T" and "S" identify tutor and student turns.

CIRCSIM-Tutor cannot have the capabilities of a human tutor. As a conversationalist, it is rather dumb. We compensate by having the computer keep control of the conversation, ending everything it says with a question or an instruction for the student. The student responds to the questions and doesn't have many opportunities to get words in edgewise.

To a first approximation the job of the input understander is to match the student's answer to CIRCSIM-Tutor's question. Often very simple, short answers suffice. Here are some short questions CIRCSIM-Tutor version 2 produces, together with typical correct student responses which we have recorded in log files:

(1) T: What is the correct value of stroke volume?
S: Increased.

(2) T: What are the determinants of cardiac output?
S: Heart rate and stroke volume.

Although we want to create more sophisticated language and tutoring strategies in version 3, resulting in richer language and occasionally more complicated questions, typical questions are not more demanding than those illustrated above.

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 of the 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 U.S. government and no official endorsement should be inferred.

To be published in 1997 Florida Artificial Intelligence Research Symposium, FLAIRS.

SIMPLE ANSWERS

A number of questions, such as those above, are *wh*-questions which admit of short answers. It is tempting to think that the input understander merely needs to match the answer to the missing constituent of the question. Unfortunately, there is often little syntactic resemblance between the question and the answer. Here are some responses to the question: "What is the correct value of (some parameter)?"

up	(adverb)
increase	(verb)
increases	
increased	(adj. or past participle)
i	(drastic but common abbreviation)
unchanged	
no change	
goes up	(phrasal verb)
went up	
it goes up	(a whole sentence)
negative	(adjective)
+	(symbol)
zero	
remains same	(curious grammar)

We have fairly extensive examples of these kinds of answers. Despite the wide syntactic variety, there are a finite number of them and little linguistic creativity. Furthermore, there are only a limited number of concepts—the above examples cover only UP, DOWN, and NO CHANGE.

We see little value in parsing most of these answers as sentences or even phrases. ("It goes up" might be a counterexample.) Instead, we have created a small grammar of short answers. This grammar is in the formalism of Lexical-Functional Grammar.

There is a category of answers which are creative phrases, meaning that they aren't fixed but are formed by linguistic processes. These are usually noun phrases. For instance students have been observed uttering "neural," "nervous system," "parasympathetic nervous system," "sympathetics," "sympathetic stimulation," "sympathetic tone" and "reflex" among other answers, all in response to the same basic question. Given such data, we would not be surprised to see "neural stimulation" or "reflex system." Thus the grammar of short answers must describe simple NPs.

Note that the eventual interpretation of the noun phrase does not necessarily depend on the head noun. The word "system" contributes nothing when used in a phrase such as "neural system," it is "neural" which carries the meaning. Our use of Lexical Functional Grammar introduces a minor awkwardness here. The result of an LFG parse is a functional representation of the input, called an *f-structure*, which produces (roughly) one sub-function per phrase. The primary meaning of each sub-function is contributed by the head word of the phrase, which in the case of a head word like "system" needs to be ignored. The lexical entries for such words are marked with a feature signaling that the interesting meaning is carried by whatever phrase modifies the head noun.

More generally, the grammar of short answers ignores most of the conventional linguistic interpretations. The f-structures resulting from parsing the utterances "up" and "increased" are nearly identical, containing the same predicate UP. These utterances are not represented as fragments reduced from more complete sentences. Instead a special feature ANS is added to the f-structure, meaning the result was parsed according to the answer grammar. A fundamental purpose of the ANS feature is to note that not all arguments of the predicate are present, or even that the main predicate is missing entirely. In LFG terms these would be ill-formed functional structures.

We occasionally observe complete sentences for which we have a more conventional sentence grammar, for example "stroke volume increases." The resulting f-structure is a conventional representation of a sentence. Clearly the input understander cannot simply note the verb "increase" (predicate UP) to find the meaning of this sentence. Suppose "stroke volume increases" had been an answer to "what is the value of cardiac output?"

Since the answer grammar doesn't cover very many whole sentences, there is no conflict in having both the answer grammar rules and the sentence grammar rules combined into one grammar, so parsing is performed once.

SPELLING ERRORS AND ABBREVIATIONS

We are using the spelling corrector written by Elmi [1994], derived in part from earlier CIRCSIM-Tutor spelling correctors. Elmi has shown that his four-way match algorithm is generally superior to several previous algorithms for comparing an unknown word to candidate words.

Since there are roughly 3000 strings in our lexicon at this time, a number which will inevitably grow, the spelling corrector often proposes several different possible corrections. We have several techniques for arbitrating, a primary one being to pick the word which most resembles the word which is the desired answer to the question.

In addition to spelling problems, one can observe in the transcripts many impromptu abbreviations. When conversing with human tutors, students often abbreviate with abandon. "Inotropic state" can become "inotropic s." "Parasympathetic" can become "parasymp" or "para." There is a stock of standard abbreviations which can be used at any time. "Cardiac output" is usually typed "co" or sometimes "c.o." The standard abbreviations are in the lexicon, and are thus not subject to spelling correction. But the impromptu abbreviations are not in the lexicon, so it is the spelling corrector which ultimately must handle them.

We have a concept ontology which we use during parsing and sentence generation for matching up nouns with the argument structure of verbs. The ontology captures the knowledge that PARAMETERs can be complements to "increase," but ANATOMICAL PARTs cannot. We plan to use this ontology, along with discourse history, to experiment with spelling correction. Suppose "ventricale" (an attested error) matches closely to both "ventricle" and "vertical" in the lexicon. We might notice that "ventricle" is an instance of ANATOMICAL PART, and other recent words were also ANATOMICAL PARTs (the error was in a discussion of anatomy), causing the input understander to prefer "ventricle."

EQUATIONS

There are a few simple equations which occur in discussions of our problem domain. Students have been observed using them in replies, even when the question did not demand their use. Here are a few examples of exchanges where students introduce equations:

(3) S: The stroke volume is the first to be affected and it will increase.
T: Next.
S: SV X HR = CO so the cardiac output will increase also. (K40-63)

(4) T: But, what are the determinants of CO?
S: HR x SV. (K42-110)

(5) T: When CO increases, does it affect the value of another variable?
S: Yes, MAP = CO * TPR. (K4-41)

In (3) the student's second utterance might more conventionally be "CO increases."

In (4) a more usual answer is "HR and SV." Sometimes a student will say "CO = HR x SV," further complicating the extraction of the desired answer.

In (5) the desired answer is "MAP."

We have similar examples in logs of students using the computer tutor, so this use of equations seems to be within the range of language which students expect the computer to understand.

Students seem to regard equations as part of their regular language, and freely intersperse them with ordinary English. Thus the input understander must incorporate these equations into the regular sentence grammar, where "=" is a verb and multiplication is a conjunction, as the following example would indicate:

(6) S: But isn't CO X TPR = MAP? (K7-100)

When a student answers a simple question with an equation, normally the input understander can extract the answer it was looking for. However every time a student answers with an unasked-for equation, there is the possibility the equation is incorrect. The student perhaps didn't intend to raise a side issue that wasn't in the tutor's plan, nevertheless the tutor might well want to correct the misconception. If the tutor is a planning engine, the effect is handled similarly to a student initiative—the tutorial planner must temporarily suspend its current tutoring plan and proceed to tutor about the unasked-for equation.

HEDGES

Hedged answers occur frequently when students are conversing with human tutors. Here are a few we have observed:

(7) T: Do you know which parameter in the prediction table determines RAP?
S: SV and CO? (K38-44)

(8) T: Ask yourself by what mechanism (or set of interactions) the MAP might be ultimately increased.
S: How about RAP. (K44-26)

(9) T: CC went down and RAP went up (we have finally decided). Which one wins?
S: RAP. I'm not sure. (K43-128)

It has long been our goal to make use of hedges in the student model and perhaps in dialogue generation. A heavily hedged correct response, for example, might conceivably deserve an explicit positive acknowledgment, even though positive acknowledgments are not always emitted.

It is hard to discern a common principle which enables the input understander to recognize hedges. We have a list of hedge phrases from our transcripts (including both "how about" and the popular question mark), which the input understander will convert to a HEDGE feature in the f-structure. It makes no practical sense to parse "I think cardiac output increases" (another attested form) into a main verb "think" with a complement sentence. Consider first that few real answers are complicated enough to contain an embedded sentence, and second that the student may also utter "cardiac output, I think."

It must be noted that students sometimes hedge by simply asking a question, perhaps with no punctuation at all, so parsing (to recognize question syntax) is useful for identifying answers hedged in this fashion.

APPROPRIATENESS OF ANSWERS

The concept ontology allows us to judge the appropriateness of a student answer which is other than the desired one. If the question was "what determines stroke volume," the answer "cardiac output" is incorrect but it is the right kind of object (a PARAMETER) in the concept ontology. The answer "the nervous system" is incorrect, and is not the right kind of object, yet it is a CAUSATIVE AGENT. Responding with the wrong kind of causative agent isn't entirely inappropriate for this question, which was eliciting which parameters cause stroke volume to change. Contrast this with responses containing words recognized by the input understander, but having no discernible relationship to the question. It is the responsibility of the input understander to identify a variety of student responses which are somewhat appropriate but not exactly right.

Whether or not the next version of CIRCSIM-Tutor can constructively use the appropriateness judgment is an open question. We have two experiments in mind. Except for "near misses" (see below), all wrong answers will trigger the same tutorial re-planning mechanism. Hume et al. [1996] seem to indicate that the tutorial tactic our human tutors pick next (in particular whether they hint or not) is dependent on the appropriateness of the student's wrong answer. If we have a rich enough set of tutoring tactics, it might be possible to use the appropriateness of a student's answer to choose among them. Another potential use of the appropriateness judgment is to pick among different kinds of acknowledgment responses that the tutor can issue. Brandle and Evens [1997] have been studying the behavior of our human tutors and developing a theory relating acknowledgments to discourse intentions.

UNEXPECTED APPROPRIATE ANSWERS

Akin to student initiatives, even though the student doesn't intend them that way, are appropriate but unexpected answers, sometimes called "near misses." Here is an example where the tutor was seeking a parameter called RAP but instead received "filling," which is determined by RAP:

(10) T: What are the parameters that determine the value of SV?
S: Filling and contractility?
T: Right, but which parameter in the table reflects filling?
S: RAP. (K44-116)

The tutorial planner is forced to treat this answer as a kind of initiative. It isn't an incorrect answer, which might cause the current tutoring tactic to fail. But it causes the tutor to modify its current plan, introducing an extra turn to get the student back to the plan.

Understanding this kind of student utterance requires quite a bit of domain knowledge. We have a "concept map," a graph of which cardiovascular parameters and objects affect which others. Judging appropriateness also requires knowing the current tutoring goal. In (10), the answer "filling" is appropriate not only because it stands between RAP and SV in the concept map, but also because it is the relationship between RAP and SV that is being tutored.

Our experience with CIRCSIM-Tutor version 2 shows us that correct recognition of these near misses is necessary. The version 2 tutor responds to a situation like in (10) by saying "filling is wrong." This is incorrect physiology—filling *is* a determinant of RAP, even if it wasn't the desired answer. No matter how well or badly CIRCSIM-Tutor performs, if it makes incorrect statements we can't let medical students use it.

MESSAGE UNDERSTANDING AND STUDENT INITIATIVES

If CIRCSIM-Tutor is to handle unconstrained student initiative input it behooves us to find the state of the art in relevant input understanding techniques. I believe the answer is to be found in an area of natural language processing called "message understanding" or (more recently) "information extraction," see for example [MUC, 1995]. Message understanding starts with real-world text. The result of the message understanding task is a set of filled-in templates, where the information slots in the templates have been determined ahead of time. The goal is to fill out those templates by whatever means necessary, extracting the information from the text.

Message understanding becomes relevant to a CIRCSIM-Tutor input understander if CIRCSIM-Tutor starts to accept student initiatives or mixed-initiative dialogue. Some of the categories we have discussed in the project are: requests for definitions, requests for explanations, and students' explanations presented to the tutor for judging [Shah and Evens 1996]. A typical simple initiative which we would categorize as a request for an explanation is "I need to go back to the start." The number of possible ways to express this fairly simple idea is quite large, but we would like to be able to respond to this kind of initiative. It is our hope that using some of the message understanding approaches will be easier and more robust than extending our input understander's grammar to a general-coverage grammar of English.

CONCLUSIONS

We observe a goodly number of interesting phenomena in the student utterances within tutoring dialogue with both humans and computers. The new version of the CIRCSIM-Tutor program must handle everything from spelling errors and impromptu abbreviations to unasked-for equations and unexpected but appropriate answers. Some of these responses resemble student initiatives, even though the student was simply answering the question. In the future we hope to have a robust mechanism for responding to genuine student initiatives as well. We have shown how the CIRCSIM-Tutor project approaches these issues.

REFERENCES

Brandle, Stefan and Martha Evens. 1997. *Acknowledgments in Tutorial Dialogue*. In preparation.

Elmi, Mohammad Ali. 1994. *A Natural Language Parser with Interleaved Spelling Correction Supporting Lexical Functional Grammar and Ill-Formed Input*, Ph.D. dissertation, Illinois Institute of Technology, Chicago.

Freedman, Reva. 1996. *Interaction of Discourse Planning, Instructional Planning and Dialogue Management in an Interactive Tutoring System*, Ph.D. dissertation, Northwestern University, Evanston.

Hume, Gregory, Joel Michael, Allen Rovick, and Martha Evens. 1996. "Hinting as a Tactic in One-on-One Tutoring," *The Journal of the Learning Sciences*, 5:1, pp. 23–47.

MUC. 1995. *Sixth Message Understanding Conference, (MUC-6): Proceedings of a Conference held in Columbia, Maryland, November 6–8, 1995*, San Francisco, CA: Morgan Kaufmann.

Shah, Farhana and Martha W. Evens. 1996. "How to Respond to Student Initiatives in Tutoring Systems," Seventh Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, Indiana.

Woo, Chong W. 1991. *Instructional Planning in an Intelligent Tutoring System: Combining Global Lesson Plans with Local Discourse Control*, Ph.D. dissertation, Illinois Institute of Technology, Chicago.