Proc. 22nd Annual Meeting of the Cognitive Science Society CogSci '2000, Philadelphia, pp. 262–267. (corr. 6/5/00).

Learning the Use of Discourse Markers in Tutorial Dialogue for an Intelligent Tutoring System

Jung Hee Kim
(janice@steve.iit.edu) IIT
Michael Glass
(michael.glass@iit.edu) IIT
Reva Freedman
(freedrk+@pitt.edu) LRDC
Martha W. Evens
(evens@iit.edu) IIT

Department of Computer Science
Illinois Institute of Technology
10 W. 31st St.
Chicago, IL 60616

Learning Research and Development Center
University of Pittsburgh
3939 O'Hara St.
Pittsburgh, PA 15260

Abstract

Usage of discourse markers in tutorial language can make the difference between stilted and natural sounding dialogue. In this paper we describe some simple rules for selection of discourse markers. These rules were derived for use in an intelligent tutoring system by applying decision-tree machine learning to human tutoring language. The fact that these selection rules operate within the environment of an intention-based planner encouraged us to derive our decision tree partly based on intention-based features. The resulting tree, when applied to the generation task, is relatively easy to understand because it can be referred to traditional intention-based linguistic explanations of discourse marker behavior.

Introduction

CIRCSIM-Tutor (CST) is a natural language-based intelligent tutoring system that engages the student in Socratic-style dialogue. The goal of the CST project is to imitate fluent simplified human tutoring language, both in the choice of tutorial dialogue strategies and in the use of language.

One feature of fluent dialogue is the use of discourse markers such as "so," "and," and "now," which often occur at structural boundaries in the discourse. Discourse markers, also known as cue words, have as many different descriptions as people describing them. In Grosz and Sidner's (1986) procedural description of discourse, discourse markers flag changes in both attentional and intentional state. In Rhetorical Structure Theory, discourse markers mark rhetorical relations between segments (Mann and Thompson, 1988). The grammar of Quirk et al. (1985, pp. 632 *ff*) subsumes most discourse markers within conjunctions. Stenstrom's (1994) manual on analyzing discourse emphasizes their use as marking boundaries of topics and digressions and describes them in concert with interpersonal "interactional signals." Schiffrin (1987) provides a detailed accounting of the behavior and purpose of eleven discourse markers without being tied to a particular theory of discourse or syntax. Schiffrin also provides an operational definition of discourse markers, giving evidence that discourse markers have functions such as aiding coherence and cohesion in text. Halliday and Hasan (1976) in their book on cohesion describe the function of quite a number of discourse markers in detail.

Recently there have been attempts to describe the behavior of discourse markers in computationally useful ways by applying methods of machine learning and corpus linguistics. Litman (1996) devised rules for distinguishing between semantic and structural uses of discourse markers in transcribed speech. In sharp distinction to the more traditional linguistic accounts, the rules are based largely on observable features such as the length of phrases, preceding and succeeding cue words, and prosodic features. Moser and Moore (1995) divided instructional dialogue into discourse segments and coded various relationships between them according to Relational Discourse Analysis, which combines Grosz and Sidner's type of analysis with Rhetorical Structure Theory. They derived rules for a number of aspects of discourse marker usage, including placement and occurrence vs. omission. Di Eugenio, Moore, and Paolucci (1997) studied the same dialogues toward similar ends. Nakano and Kato (1999) studied Japanese instructional dialogue, using machine learning to derive rules for occurrence of three categories of discourse markers. They divided their text into segments in the same manner as RST, but also coded the instructional goals for each segment in addition to coding the kinds of features used in previous studies.

The addition of instructional goals in Nakano and Kato's study is important to the CIRCSIM-Tutor project, and should be encouraging from the standpoint of trying to generate (as opposed to analyze) instructional dialogue. One reason is that instructional goals proved to be explanatory. A common feature of the machine learning studies is that the text is coded for a large number of features, of which only a few are incorporated by the machine learning process into the eventual rules or decision tree. In Nakano and Kato's study instructional goals were so incorporated, meaning that they were more explanatory than many of the other features. This is congruent with non-corpus-based linguistic theories that explain discourse markers in terms of the speaker's intentions.

The speaker's intentions are rarely explicit in text; for purposes of analysis intentions are divined by coders. However when the machine tutor is generating dialogue, the machine speaker's "intentions," i.e. the tutorial goals, can be given in the form of planning goals, see for example (Young, Moore & Pollack, 1994). Nakano and Kato have shown that having the tutorial goal structure in hand can potentially lead to better discourse marker selection.

In this paper we use attribute-based machine learning of decision trees, specifically the C4.5 algorithm (Quinlan, 1993), to investigate discourse marker selection. We make use of both structural features and aspects of the sequence of tutorial goals—the "intention" of the machine tutor. Although we learn rules from transcripts of human dialogues, we concentrate on features that are available within the CIRCSIM-Tutor generation environment.

The machine tutor does not reason about rhetorical relations such as are usually used to explain discourse markers. Instead it has planning goals that produce schemata containing patterns of dialogue. These schemata define the dialogue segments. Rhetorical relations are implicit in the patterns, so it is possible to relate goal-structure explanations of discourse markers to the rhetorical relation-based theories.

The Experiment

We recorded the features surrounding instances of discourse markers in human tutorial dialogue, then derived a decision tree to predict discourse marker selection.

The users of CIRCSIM-Tutor are medical students in a first-year physiology class studying the reflex control of blood pressure. Students are required to predict the changes in a set of physiological variables, after which the tutor endeavors to elicit corrected predictions via Socratic-style dialogue, asking questions and giving hints. CST's conversation can be largely segmented into the correction of individual variables.

The CIRCSIM-Tutor project has transcripts of one- and two-hour keyboard-to-keyboard tutoring sessions between physiology professors and medical students. Our construction of the computer tutor's planning operators and tutorial language is informed by these transcripts. The transcripts were previously marked up with tutorial goals and language phenomena for this purpose (Kim, Freedman & Evens, 1998a, b; Freedman et al., 1998; Zhou et al., 1999). Tutorial goals consist of global goals for tutoring and local goals for maintaining coherence of dialogues. The global goals used in this study are hierarchically arranged into *method* and *topic* levels. A *method* goal describes one way to remediate a student's incorrectly predicted physiological variable. Within one method, a sequence of *topic* goals describes individual concepts to be expressed. A topic can be expressed by either telling the information to the student or eliciting it from the student. A typical dialogue pattern for the correction of one individual variable is as illustrated in Figure 1. The sequence of tutorial goals is as follows:

- The variable to be corrected is introduced into the conversation.
- Various topic goals are realized by telling them to the student or eliciting them from the student.
- The corrected prediction is elicited from the student.

The discourse markers we study in this paper occur at the boundaries between topic goals, as shown in italics in Figure 1. We are concerned with the selection of these discourse markers in human tutorial dialogues in order to generate them correctly. Placement of discourse markers is not an issue, we ignore discourse markers which occur elsewhere.

It will be noted that in our dialogues the junctures between topic goals do not always coincide with the turn boundaries; in fact in our illustration one topic is spread among three turns and one turn encompasses parts of three topics. One typical tutor turn contains:

- An optional acknowledgment of the student's answer
- Possibly an elaboration on that answer
- Possibly some new information
- A question or instruction to the student

(Freedman & Evens, 1996)

The context of a discourse marker therefore includes not only the structure of topic goals, but also information from the turn structure. Preceding the first discourse marker in a tutor's turn is a possible tutor's acknowledgment to the student and possibly some elaboration. Furthermore there is the student's immediately preceding turn, which usually consists of the answer to the tutor's previous question. Some examples of these features, including our characterization of the correctness of the student's answer, are also annotated in Figure 1.

The human transcripts also contain dialogue that is too complex for us to mark up according to our goal hierarchy and is therefore excluded from our sample.

We further restricted ourselves to exchanges where the student gave answers that were correct or "near misses." A near miss is a student answer that is true but not expected, and can be repaired without contradicting the student (Zhou et al., 1999). In the dialogue in Figure 1, the tutor repaired

the student's overly specific answer by echoing back the more general answer. Sometimes the tutor temporarily suspends the current topic goal and interpolates a tutoring schema to repair the unexpected answer. In that case the goal hierarchy would show an inner sequence of topic goals devoted to remediating one outer topic. These instances are included in our sample. The tutor's responses to incorrect student answers (as opposed to near misses) are too varied for us to obtain any regularities in discourse marker usage, so we excluded them.

We extracted instances of the discourse markers "and," "so," and "now" because these are the most frequently used ones in our transcripts. Each instance consists of the context around one discourse marker coinciding with a topic change, coded for the following five attributes:

- Category of the student's answer preceding the marked topic boundary: correct, near miss, or N/A. The N/A case occurs when the tutor covers several topics within one turn, so the topic preceding the discourse marker does not contain a student answer.
- Presence or absence of acknowledgment preceding the topic boundary: ack, no-ack, N/A.
- Discourse marker: "and," "now," "so."
- Position within the sequence of topic goals of the topic following the discourse marker: introduce, initial, middle, or final.
- Presentation of the topic following the discourse marker: inform or elicit.

Thus the sentence "and the reflex hasn't started to operate yet" from turn 3 of Figure 1 is coded as:

- Student's answer category = "near miss"
- Acknowledgement = "present"
- Discourse marker = "and"
- Position in sequence = "middle"
- Type of presentation = "inform"

We supplied 60 cases of these feature-annotated discourse marker occurrences to the C4.5 machine learning program. It produced the following rules for selection of the discourse marker:

- If the topic position is introduce then use "now"
- If the topic position is middle then use "and"
- If the topic position is final then use "so"
- If the topic position is initial
  and if the presentation is inform then use "so"
  else {presentation is elicit} use "and"

These rules misclassified 8 of the 60 cases, for an error rate of 13.3%.

These rules describe our expert tutors' linguistic behavior, predicting which discourse marker will be selected in certain contexts. We start with this description in order to produce rules for text generation.

Discussion

Most of the predictions of the derived rules can be explained by existing discourse marker theories. The "now" on the introduction topic is consistent with the explanation by Grosz and Sidner (1986) of marking an attentional change, creating a new focus space of salient objects and topics. Schiffrin (1987, p. 230) says "...'now' marks a speaker's progression through discourse time by displaying attention to an upcoming idea unit." In fact, this reading of "now" explains some of the cases of "now" that are misclassified by the derived rules. These are cases where the tutor does not explicitly utter an introduce topic at the beginning of the segment, with the result that the attention-shifting "now" is attached to the initial topic. Here is one example:

> Now, what two parameters in the prediction table together determine the value of SV?

Athough the derived rules misclassify our marked-up transcripts in these cases, for the purpose of generating sentences in the machine tutor this is a useful discovery. The intention to shift tutoring to a new variable is available in CIRCSIM-Tutor's tutorial goal structure, even if not always expressed in text, so the text generator can plausibly know to emit "now."

Most of the remaining predictions of the derived rules can be explained by existing discourse marker theory. Shiffrin (1987) and Halliday and Hasan (1976) and Quirk et al. (1985, p. 638) all describe "so" as indicating a result. In our derived rules, the "so" attached to the final topic is used in this fashion. The final sentence of turn 3 in Figure 1 illustrates this point.

When the rules predict "so" attached to the initial topic it has a different role. It is found in what we call the *present-anomaly* tutoring method used to point out the inconsistent appearance of reported facts, viz:

> So, in DR heart rate is up, cardiac output is up, but stroke volume is down. How is this possible?

This "so" is explained by Halliday and Hasan as "a statement about the speaker's reasoning process" meaning it is logical to be having this thought right now.

The discourse marker "and" usually occurs on medial topics to "coordinate and continue" the topics (Schiffrin, 1987, p. 152), and needs no explaining. The discourse marker "and" occurring on the initial topic seems anomalous, but it occurs in the context of a tutorial schema we call *move forward*. This schema attempts to persuade the student of the correct value of a new physiological variable based on the result of the immediately preceding discussion of a different variable. Here is an example:

Tu: ...That being the case, what will happen to right atrial pressure in this situation?
St: Increase.
Tu: *And* if right atrial pressure increases, what would happen to stroke volume?

In this example, the final topic in the first segment occured when the student produced the correct value for right atrial pressure. The tutor skipped introducing the next variable, stroke volume, and proceeded directly to the initial topic of the tutoring schema for its correction, which moves forward in causal physiological reasoning from the final topic in the preceding segment. In this case "and" is warranted, it would seem that "so" would be equally appropriate. This is another instance where the CIRCSIM-Tutor text generator makes use of the discourse goal being processed. Even though tutoring of a new variable usually starts with the discourse marker "now," when the new variable is taught by the *move forward* method goal then the generator emits "and" instead.

Except for the initial discourse marker (usually "now") at the beginning of a tutoring method schema, it is possible to apply to our own data Di Eugenio et al.'s (1997) discoveries relating rhetorical structure to discourse marker occurrence. Although we did not perform any rhetorical structure analysis on our texts, most of our method schemas fit one of their patterns, as described next.

Here is an idealized realization of a typical CIRCSIM-Tutor method schema for teaching a variable, called *tutoring via determinants*:

Tu: What are the determinants of cardiac output?
St: Heart rate and stroke volume.
Tu: And what is the relation of stroke volume to cardiac output?
St: Direct.
Tu: And we have already seen that heart rate is unchanged. So what happens to cardiac output?
St: It goes up.

In order to analyze this in terms of rhetorical relations, we write down all the propositions in the sequence they occur as if it were a monologue, thereby exposing the argument in simplest form. Since the intention of each of the tutor's questions is to cause the student to believe the corresponding assertion, we think this is a reasonable model.

a) The determinants of cardiac output are heart rate and stroke volume,
b) And stroke volume affects cardiac output directly,
c) And heart rate is unchanged,
d) So cardiac output goes up.

In the terms of Relational Discourse Analysis, proposition d) is the *core* while a), b), and c) are *contributors*. The intentional relationship between each contributor and the core is *convince*. In fact, most of our methods have the same structure: the core is the last statement, where the value of the variable is finally understood, and the contributors all argue for the truth of the core. In (Di Eugenio et al., 1997) these relations are all analyzed in the "core2" class, meaning that the core follows the contributor in the text. Their decision tree on discourse marker occurrence yields a simple answer for these cases: the discourse marker should ordinarily appear.

Conclusions

We have applied decision tree learning to transcripts of expert tutors in order to learn rules that predict discourse marker selection. Our purpose in this endeavor is not to find rules for analyzing texts, but to produce rules for text generation in CIRCSIM-Tutor. Discourse marker usage has traditionally been explained partly in terms of the intention of the speaker and partly in terms of the rhetorical structure of the text. Neither is explicit in transcripts of discourse, but must be imputed by researchers before analyses of discourse markers can proceed. Recent work in using machine learning to explain discourse marker usage has thus shied away from using intention-based explanations.

However within the context of the machine tutor the generation algorithm has access to the speaker's intentions. In CIRCSIM-Tutor these intentions are the pedagogical goals. The structure of these goals implies the rhetorical structure of the text to be generated. So without explicit reasoning in the rhetorical terms that usually explain discourse markers, simply examining the current goals enables the text generator to select the correct discourse marker.

Our machine-derived decision tree analysis of discourse marker selection is quite successful. The features that drove the machine learning process included the same pedagogical goal analysis as is used by the machine tutor. The decision tree that results was examined by hand; where it incorrectly predicts observed data the decisions can be enhanced by applying traditional linguistic explanations. The fact that this decision tree is intention-based enables us to correlate it to existing linguistic descriptions of discourse marker usage.

Acknowledgments

Joel A. Michael and Allen A. Rovick, professors of physiology at Rush Medical College, are responsible for the pedagogical and domain knowledge in CIRCSIM-Tutor, and served as expert tutors for the transcripts. Yujian Zhou helped bring machine learning to the CIRCSIM-Tutor project, and has been helpful in all endeavors.

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

Di Eugenio, Barbara, Johanna D. Moore, and Massimo Paolucci. (1997). Learning features that predict cue usage. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL '97)*, Madrid, Spain, pp. 80–87.

Freedman, Reva and Martha W. Evens. (1996). Generating and Revising Hierarchical Multi-Turn Text Plans in an ITS. In C. Frasson, G. Gauthier and A. Lesgold, eds., *Intelligent Tutoring Systems: Third International Conference (ITS '96)*, Montreal. Berlin: Springer-Verlag. Lecture Notes in Computer Science, no. 1086.

Freedman, Reva, Yujian Zhou, Michael Glass, Jung Hee Kim, and Martha W. Evens. (1998). Using rule induction to assist in rule construction for a natural-language based intelligent tutoring system. *Proceedings of Twentieth Annual Conference of the Cognitive Science Society*, Madison, WI, pp. 362–367.

Grosz, Barbara J. and Candace L. Sidner. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3): 175–204. Also published as a technical report in 1986 with the same title: Technical Report no. 380 from the Center for the Study of Reading, University of Illinois at Urbana-Champaign.

Halliday, M. A. K. and Ruqaiya Hasan. (1976). *Cohesion in English*. London: Longman.

Kim, Jung Hee, Reva Freedman, and Martha W. Evens. (1998a). Relationship between tutorial goals and sentence structure in a corpus of tutoring transcripts. *Proceedings of Ninth Midwest AI and Cognitive Science Conference(MAICS '98)*, Dayton, OH, pp. 124–131.

Kim, Jung Hee, Reva Freedman, and Martha W. Evens. (1998b). Responding to unexpected student utterances in CIRCSIM-Tutor v. 3: Analysis of transcripts. *Proceedings of the Eleventh Florida AI Research Symposium (FLAIRS '98)*, Sanibel Island, FL, pp. 153–157.

Litman, Diane J. (1996). Cue phrase classification using machine learning. *Journal of Artificial Intelligence Research*, 5: 53–94.

Mann, William C. and Sandra A. Thompson. (1988). Rhetorical Structure Theory: Towards a functional theory of text organization. *Text*, 8(3): 243–281.

Moser, Megan and Johanna D. Moore. (1995). Investigating cue selection and placement in tutorial discourse. *Proceedings of the 33th Annual Meeting of the Association for Computational Linguistics (ACL '95)*, Boston, MA, pp. 130–135.

Nakano, Yukiko I. and Tsuneaki Kato. (1999). Cue phrase selection in instruction dialogue using machine learning. *Discourse Relations and Discourse Markers, Proceedings of the Workshop of COLING-ACL '98,* Montreal, pp. 100–106.

Quinlan, J. Ross. (1993). *C4.5: Programs for Machine Learning*, San Mateo, CA: Morgan Kaufmann.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Schiffrin, Deborah. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Stenström, Anna-Brita. (1994). *An Introduction to Spoken Interaction*. London: Longman.

Young, R. Michael, Johanna D. Moore and Martha E. Pollack. (1994). Towards a principled representation of discourse plans. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Atlanta.

Zhou, Yujian, Reva Freedman, Michael Glass, Joel A. Michael, Allen A. Rovick, Martha W. Evens. (1999). What should the tutor do when the student cannot answer a question? *Proceedings of the Twelfth International Florida AI Research Symposium (FLAIRS '99)*, Orlando, pp. 187–191.

Turn	Text	Global Tutoring Goal	Other Features
1. Tu:	*Now* let's look at your prediction for TPR.	Inform introduce variable	Discourse Marker = *Now*
	Can you tell me how it is controlled?	Elicit initial topic	
2. St:	Parasympathetics		Answer Category = Near Miss
3. Tu:	Correct, TPR is neurally controlled.		Acknowledgment = *Correct*
	And the reflex hasn't started to operate yet.	Inform middle topic	Discourse Marker = *And*
	So what is the value of TPR?	Elicit final topic	Discourse Marker = *So*
4. St:	Unchanged		Answer Category = Correct
5. Tu:	Great!		Acknowledgment = *Great*
	What other variables are neurally controlled?	Introduce next variables.	

Figure 1. Annotated Tutorial Dialogue for Correcting One Variable.