Pilot Test of a Natural-Language Tutoring System for Physics That Simulates the Highly Interactive Nature of Human Tutoring

Sandra Katz[1], Patricia Albacete[1], Michael J. Ford[2], Pamela Jordan[1],
Michael Lipschultz[3], Diane Litman[1,3], Scott Silliman[1], and Christine Wilson[1]

[1] Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA, USA
{katz,palbacet,pjordan,scotts,clwilson}@pitt.edu
[2] School of Education, University of Pittsburgh, Pittsburgh PA, USA
mjford@pitt.edu
[3] Dept. of Computer Science, University of Pittsburgh, Pittsburgh PA, USA
{dlitman,lipschultz}@cs.pitt.edu

Abstract. This poster describes Rimac, a natural-language tutoring system that engages students in dialogues that address physics concepts and principles, after students have solved quantitative physics problems. We summarize our approach to deriving decision rules that simulate the highly interactive nature of human tutoring, and describe a pilot test that compares two versions of Rimac: an experimental version that deliberately executes these decision rules within a Knowledge Construction Dialogue (KCD) framework, and a control KCD system that does not intentionally execute these rules.

Keywords: Natural-language dialogue, human tutoring, interaction hypothesis.

1 Introduction

Much research has been devoted to identifying features of tutorial dialogue that can explain its effectiveness [1], so that these features can be simulated in natural-language tutoring systems. One hypothesis is that the highly interactive nature of tutoring itself promotes learning—that is, the "Interaction Hypothesis" [2]. Although this hypothesis is supported by some research, several studies indicate that it is underspecified. That is, it is not how much interaction takes place during tutoring that is important, nor the granularity of interaction—for example, whether the student and tutor discuss a step towards solving a problem, or the sub-steps that lead to that step. Instead, what matters most is how *well* the interaction is carried out (e.g., [3]).

This refinement of the Interaction Hypothesis raises the question, *which linguistic mechanisms support learning from human tutoring?* We address this question by identifying co-constructed discourse relations in tutorial dialogues whose frequency predicts learning; specifying the context in which these relations occur; using this knowledge to formulate decision rules to guide automated tutorial dialogues; implementing these rules in a natural-language tutoring system; and testing the effectiveness of this decision rule-based system, relative to a system that does not intentionally execute these rules [4].

Table 1. Examples of co-constructed discourse relations during physics tutoring

Bi-directional relations and definitions (S=Speaker)	Examples (T=tutor; S=student)
Whole:Part (**Part:Whole**) S2 names a part of an object that S1 refers to; or S1 names a part of an object named in S2.	**S:** acceleration would be plus **T:** right, the x-component of the acceleration would be plus
Process:Step (Step:Process) S2 presents a step that follows from the process or line of reasoning described by S1; or S2 describes the line of reasoning that leads to the step described by S1.	**S:** the acceleration is zero **T:** so then m*a=0=Fnet=T-W and hence T=W.
Condition:Situation (Situation:Condition) S1 presents a condition or set of circumstances, and S2 states the situation that stems from those conditions; or, S1 presents a situation and S2 states the conditions or circumstances that explain it.	**T:** when do kinematics equations apply? **S:** when the acceleration is constant

We used Rhetorical Structure Theory [5] to identify and tag co-constructed discourse relations in a large corpus of instructional dialogues between human physics tutors and students, via typed interaction ([6], study 2). A sample of these relations are defined and illustrated in Table 1. Any relation can be delivered didactically, by the student or tutor, instead of interactively. For example, the co-constructed Condition:Situation (conditional) relation shown in Table 1 could have been stated didactically by the tutor as, "Kinematics equations apply when the acceleration is constant." We focused on the potential relationship between *co-constructed* discourse relations and learning because these relations operationalize vague notions such as "interactivity" and "cooperative execution" during tutoring (e.g., [2], p.199).

We found that the frequency of several types of co-constructed relations in the tagged corpus predicted learning gains from pretest to posttest. Moreover, the types of co-constructed discourse relations that predict learning vary based on students' ability level. These correlational analyses of co-constructed discourse relations and learning, and the decision rules that stem from them, are described in detail in [4]. In this poster, we: (1) illustrate these decision rules, which are implemented in Rimac, a natural-language tutoring system that guides reflective dialogues about the concepts associated with quantitative physics problems, and (2) describe the design of a pilot evaluation of Rimac that we are currently conducting, and planned analyses.

2   Methods

2.1   Deriving Decision Rules to Guide "Highly Interactive" Dialogues

We conducted correlational analyses between the frequency of discourse relation types (Table 1) and three measures of student learning: overall gain score from pretest to posttest, gain score on qualitative test items, and gain score on quantitative items.

We divided students into ability groups to investigate whether better-prepared students (high pretesters) might benefit from co-constructing different types of discourse relations with their tutor than less well-prepared students (low pretesters). These analyses, coupled with an analysis of the discourse context in which potentially effective relations occur, enabled us to formulate nine decision rules to drive "highly interactive" tutorial dialogues [4]. Two examples of these decision rules are:

Rule 1: **When the student provides a step in a line of reasoning, the tutor may provide the missing steps, rather than ask about each step individually.** This decision rule stems from several correlations involving the Step:Process relation. For example, for the set of students taken as a whole, the frequency of tutor extensions of the student's line of reasoning predicted overall gain [$r(14)=.65$, $p<.01$]. Tutors typically did this when the student answered a question correctly but not completely, or had difficulty figuring out the next step in a solution or discussed line of reasoning.

Rule 2: **If the student answers a question incorrectly, if possible show why it is incorrect by stating the conditions under which it *would* be correct.** This rule is mainly motivated by a correlation between the frequency of co-constructed conditional relations and qualitative gains among low pretest students [$r(7)=.68$, $p< .05$].

2.2   Pilot Evaluation of Rimac

We implemented the nine derived decision rules in the experimental version of Rimac, using a Knowledge Construction Dialogue (KCD) framework (e.g., [7]), but not in a control version whose dialogues are otherwise the same in content and structure. For example, in the following dialogue excerpt, the computer tutor (T) applies Rule 2 (boldfaced segment) in the experimental (decision-rule) driven version:

> **T:** When an object is slowing down, how does the final velocity (vf) compare to the initial velocity (vi) for any interval of time? (smaller, larger, etc.)?
> **S:** Larger
> **T: If the object is speeding up then its final velocity is larger than its initial velocity.** But when an object is slowing down its velocity is getting smaller all the time. So for any interval of time the final velocity is smaller than the initial velocity.

The standard KCD dialogue excerpt is the same, except for the omission of the boldfaced segment and the connector "but;" that is, it simply corrects the student. The research platform is illustrated further at https://sites.google.com/site/rimacdemo [8].

   Data collection for this pilot study is in progress, in physics classes at six high schools in the mid-western USA.   Approximately 250 students are participating in the field trials, which take place during two physics lab periods, each lasting approximately 1.5 hrs. in conjunction with the course units on kinematics and dynamics.

3   Plans for Data Analysis

We will verify that the two conditions did differ significantly, in terms of frequency of rule firings and other linguistic indicators of a high level of "interactivity." We will then determine if there is a significant difference in the amount of learning gains (from pretest to posttest) between conditions, and investigate whether there is an interaction between student ability and dialogue condition: for example, is the highly interactive version more helpful for low ability students (as measured by pretest and SAT scores), whereas the standard version better supports high ability students?

References

1. Bloom, B.S.: The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher 13, 4–16 (1984)
2. Van Lehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist 46(4), 197–221 (2011)
3. Chi, M., VanLehn, K., Litman, D., Jordan, P.: An Evaluation of Pedagogical Tutorial Tactics for a Natural Language Tutoring System: A Reinforcement Learning Approach. International Journal of Artificial Intelligence in Education 21, 83–113 (2011)
4. Katz, S., Albacete, P.: A Tutoring System that Simulates the Highly Interactive Nature of Human Tutoring. Educational Psychology (Special Issue on Advanced Learning Technologies) (in press)
5. Mann, W.C., Thompson, S.: Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text 8(3), 243–281 (1988)
6. Katz, S., Allbritton, D., Connelly, J.: Going Beyond the Problem Given: How Human Tutors Use Post-Solution Discussions to Support Transfer. International Journal of Artificial Intelligence and Education 13(1), 79–116 (2003)
7. Rosé, C., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., Weinstein, A.: Interactive Conceptual Tutoring in Atlas-Andes. In: Moore, J.D., Redfield, C.L., Johnson, W.L. (eds.) Artificial Intelligence in Education, pp. 256–266. IOS Press, Amsterdam (2001)
8. Jordan, P., Albacete, P., Ford, M.J., Katz, S., Lipschultz, M., Litman, D., Silliman, S., Wilson, C.: The Rimac Tutor - A Simulation of the Highly Interactive Nature of Human Tutorial Dialogue. In: Lane, C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 928–929. Springer, Heidelberg (2013)