Reformulating student contributions in tutorial dialogue*

Pamela Jordan[1]
pjordan@pitt.edu

Sandra Katz[1]
katz@pitt.edu

Patricia Albacete[1]
palbacet@pitt.edu

Michael Ford[2]
mjford@pitt.edu

Christine Wilson[1]
clwilson@pitt.edu

University of Pittsburgh
Learning Research & Development Center[1]
School of Education[2]
Pittsburgh PA 15260, USA

Abstract

While some recent work in tutorial dialogue has touched upon tutor reformulations of student contributions, there has not yet been an attempt to characterize the intentions of reformulations in this educational context nor an attempt to determine which types of reformulation actually contribute to student learning. In this paper we take an initial look at tutor reformulations of student contributions in naturalistic tutorial dialogue in order to characterize the range of pedagogical intentions that may be associated with these reformulations. We further outline our plans for implementing reformulation in our tutorial dialogue system, Rimac, which engages high school physics students in post problem solving reflective discussions. By implementing reformulations in a tutorial dialogue system we can begin to test their impact on student learning in a more controlled way in addition to testing whether our approximation of reformulation is adequate.

1 Introduction

In the current study of tutorial dialogue we describe here, we seek to identify the most pedagogically valuable ways in which a tutor incorporates a student's contribution into his turn so that we can implement these in a tutorial dialogue system. In educational research, two teaching techniques that have been shown to benefit students, Accountable Talk (O'Connor and Michaels, 1993) and Questioning the Author (Beck et al., 1996), both train teachers to make use of a number of discussion moves that react to student contributions. One such move that is shared by both teaching techniques is *revoicing*. Revoicing is characterized as a reformulation of what the student said with the intention of expressing it in a way that most of the student's fellow classmates will be able to make sense of it and elaborate upon it. In the case of Accountable Talk it also includes the intent that the teacher attempt to relinquish authority on the topic under discussion. This is done by not evaluating the student contribution and instead inviting the student to assess the teacher's reformulation. In tutorial dialogue, the pedagogical intention of revoicing cannot be exactly the same. However, a reformulation that invites the student to assess it may retain some of the benefits of classroom revoicing. This is something we intend to test as part of our research. A step we are taking towards such a test is to look at what reformulations appear in tutorial dialogue and then attempt to characterize the tutor intentions that may be associated with them.

In some applied contexts, such as second language learning, reformulations are more narrowly defined as using different words while keeping the content semantically equivalent. However, research in pragmatics takes a broader view of reformulation. In a corpus study of lectures that examined reformulation markers such as "in other words," "that is" and "i.e." and also endeavored to consolidate the findings from previous linguistics studies, the range of

*The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100163 to the University of Pittsburgh. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

INLG 2012 Proceedings of the 7th International Natural Language Generation Conference, pages 95–99,
Utica, May 2012.

intentions identified include, among others, definition, denomination, specification, explanation, correction and consequence (Murillo, 2008). In our preliminary characterization of reformulations in naturalistic tutorial dialogue, we will use this broader definition and will test whether a tutor contribution is a reformulation of what the student said by checking the felicity of inserted reformulation markers such as "in other words."

Two recent studies of tutorial dialogue specifically recognize revoicing. The first study (Chi and Roy, 2010) examines face to face naturalistic tutorial dialogue in which a tutor is helping a student work through a physics problem. They suggest that when the tutor repeats part of what the student said, it is often done with the intention of providing positive feedback for correct answers and call this revoicing as with the excerpt below which is repeated from (Chi and Roy, 2010) .

> S: First the gravity is pulling down
> T: Pulling it down. [Tutor revoiced.]
> S: Weight is..the mass times..acceleration due to gravity and that's force.
> T: Right. Right.
> S: Ok.
> T: So weight is the force. [Tutor revoiced.]

Given the limited context of these transcribed excerpts it is difficult to argue that these are revoicings in the sense of Accountable Talk (AT). There are no implicit or explicit invitations, such as a question mark, to assess the tutor's contributions.

While it is possible in the first example that the tutor understood the student to be making a generic statement and was adding "it" to apply it to the particular problem under discussion, it is also possible they have the shared goal of identifying and summing all the forces on a particular object and the tutor is just acknowledging understanding.

The second example seems to draw attention to what is most important in what the student just said. In AT and Questioning the Author (QtA), this type of move is called *marking* instead of *revoicing*. A marking is a reformulation that emphasizes what is most important in what the student said and attempts to direct the student to focus his/her continued discussion on the reformulation.

Although neither of these examples are revoicings in the sense of AT and the first seems more like a repetition to acknowledge rather than reformulate, both are still important to consider for tutorial dialogue. They may help lessen the student's cognitive load (Walker, 1996) by drawing attention to what is most important in what the student said (Becker et al., 2011).

The other recent study of tutorial dialogue that considers revoicing collected a corpus using human tutors who were trained to use QtA and who fill in for a conversational virtual tutor in a science education system (Becker et al., 2011). This corpus has been annotated along multiple dimensions. Two discussion moves from QtA, revoicing and marking, which are noted to be frequent in this corpus, are included in the dialogue act dimension along with other more general speech acts. However, there is no stated goal to annotate other reformulations. So we do not know what other intentions associated with reformulations may appear in the corpus.

In addition, the authors' description of revoicing differs from that used in AT. Here, it is a reformulation that is meant to help a student who is struggling with a particular concept. As shown in the annotated example of revoicing repeated below from (Becker et al., 2011), authority is not relinquished and the student is not invited to assess the reformulation.

> S33: well when you scrub the the paperclip to the magnet the paperclip is starting to be a magnet [Answer/Describe/Process]
> T34: very good, so if the magnet gets close to the paperclip it picks it up [Feedback/Positive/None, Revoice/None/None]

A range of reformulations are recognized in other work on tutorial dialogue and have been incorporated into tutorial dialogue systems. In AutoTutor (Person et al., 2003), elaboration and summary involve reformulation. In Circsim-Tutor (Freedman, 2000), student answers that are close to correct except for terminology trigger a reformulation. Finally, in Beetle II (Dzikovska et al., 2008), restatements of correct and near correct answers involve reformulations. In our work we wish to identify a more comprehensive set of reformulation types and intentions and determine which of these types are most beneficial to emulate.

In this paper we examine a corpus of naturalistic human tutorial dialogues for tutor reformulations. We further outline our plans for implementing revoicing and reformulation in our tutorial dialogue system, Rimac (Katz et al., 2011), which engages high school physics students in post problem solving reflective discussions. By implementing reformulations and revoicings we can begin to test their impact on student learning in a more controlled way in addition to testing whether our approximations of them are adequate.

First, we will describe the corpus of human tutorial dialogues we are analyzing and then we will present examples of some of the reformulations we have found in the corpus and speculate upon possible tutor intentions for these reformulations. We will then outline our plans for implementing certain types of reformulation by first describing the current tutorial dialogue system and the planned modifications for implementing tutor reformulations.

2 The Corpus

The corpus of human tutorial dialogues we are analyzing was collected during a study (Katz et al., 2003) on the effectiveness of reflection questions after a physics problem-solving session with the Andes physics tutoring system (VanLehn et al., 2005). The tutors in this corpus were graduate teaching assistants who had experience in tutoring physics. The students were recruited from introductory undergraduate physics courses.

The students first solved a problem using the Andes system and afterwards they were presented with a deep-reasoning reflection question which they needed to answer. After typing their answer, they then engaged in a typed dialogue with a human tutor to follow up on their answer. This dialogue continued until the tutor was satisfied that the student understood the correct answer. Three to eight reflection questions were asked per problem solved in Andes. There were 12 Andes problems in all.

3 Characterizing Reformulations in Reflective Tutorial Dialogue

As part of our analysis of the corpus described in the previous section, we have been annotating cases of repetition and reformulation across immediately adjacent tutor-student and student-tutor turns (Katz et al., 2011). While this effort is still ongoing and we cannot yet fully characterize the reformulations found, we can show examples of some of the reformulations we have identified and speculate upon what the tutor's intentions may have been. Our goal in this section is to show the variety of intentions one can attribute to these reformulations. Due to space limitations we cannot include examples of the full range of intentions we have found.

The first example, shown below, reformulates what the student said (in italics) by using terminology that is typical to mathematics/physics (in bold). Arguably, "I would call that" may act as a reformulation marker in this example. At the end of a reformulation, we list in square brackets the pragmatics labels we believe best characterize the reformulation.

> T: what direction (in words) is the displacement?
> S: *downwards/towards the negative y-axis*
> T: right: **I would call that the -y direction** [denomination]

The next example, shown below, reformulates what the student said in terms of a more fully specified definition. Inserting "in other words" after "Right" seems felicitous.

> T: What is speed?
> S: *it is velocity without direction*
> T: Right, **The (instantaneous) speed is the magnitude of the (instantaneous) velocity**. [specification/definition]

The next example, shown below, reformulates some of what the student said so that it is correct. Here we can insert the marker "you mean" in front of "the mass and acceleration are related to forces" and arguably "as you point out" could be serving as an explicit reformulation marker. In this case the tutor seems to be correcting an implied "equated to" to "related to."

> S: *the mass and the acceleration push the man into the airbag*
> S: *so aren't they considered forces?*
> T: **the mass and acceleration are related to**

forces as you point out, but in Newtonian mechanics are not considered forces. [correction]

And finally, the example shown below is a reformulation that is a revoicing. In this case the student may be struggling to explain but seems to have a correct conceptual understanding. The tutor attempts to summarize in a clearer way what he thinks the student meant and invites a student assessment with "I think I see what you mean" and the question mark.

S: no *gravity is no effecting x directly, but if it did not effect y, it would go on forever, and x would countinue to grow as well, but since y has a bound, so does the x*
T: **I think I see what you mean. That when gravity pulls the ball back to the earth, that the earth then affects the horizontal motion (by direct contact), which wouldn't have happened without gravity?** [summary]
S: gravity is needed to bring y back to 0 so that the d_x comp is = d

4 The Rimac Tutorial Dialogue System

To understand how we propose to implement reformulations, we must begin with a high level description of the current Rimac system. To build Rimac, we used the TuTalk (Jordan et al., 2007) natural language (NL) tutorial dialogue toolkit. This toolkit enables system developers to focus on developing the content to be presented to students and rapidly developing an end-to-end system for conducting experiments that determine what content and presentation is most pedagogically effective. Tutorial dialogue system developers can gradually transition towards a more principled dialogue system as questions of pedagogical effectiveness are answered, since core modules such as NL understanding and generation are designed to be replaced or supplemented as needed.

The simplest dialogue one can write using this toolkit can be represented as a finite state machine. Each state represents a tutor turn. The arcs leaving the state correspond to all classifications of a student's response turn. When creating a state, the author enters the NL text for a tutor's turn and enters the NL text that defines several classes of student responses as transition arcs, and indicates which state each arc leads to. An arc can also push to another finite state network.

In this toolkit, the NL text associated with a state or an arc is represented by a concept definition. In the simplest case, a concept is a set of NL phrases. For instance, the set for a concept labelled NEG-ACK might be "Not quite," "Well, not exactly," "No." When a student turn is received, the dialogue manager sends a request to the understanding module to determine what concepts it best represents and determines transitions on the basis of the concept labels returned. Likewise when a concept is to be expressed, the dialogue manager asks the generation module to determine how to best express it in NL.

In the next section we outline an initial approach for rapidly testing which reformulations matter to student learning.

5 Implementing Reformulation in Rimac and Future Work

In our preliminary approach for emulating some of the types of reformulation we have found in the corpus, if there is a more preferred phrasing for a matched concept, regardless of whether the student's response is considered correct or incorrect, then a reformulation with the preferred phrasing is presented to the student. How the reformulation is presented depends on the quality or confidence of the concept match. If the student turn is a poor match for a concept, but it is the best available match then the system will revoice the student response; e.g.:

S: the *distance of the hailstone*
T: **Are you saying 'distance the hailstone travels'**?

In this example, we assume that "distance of the hailstone" is a poor match for the expected correct concept and it does not match any other concepts that represent an incorrect response. Further, we assume that the author indicated "distance the hailstone travels" as one of the most highly ranked phrases for expressing the correct concept. So the tutor's response revoices using the preferred phrasing. In this example, the tutor's question invites the student to confirm the system's reformulation of her response or try again.

However, if the student's response is a high quality match for a concept but the author ranked this

phrasing as a less preferred way of expressing the concept, then the system will reformulate using the most preferred phrasing for the concept; e.g.:

> S: *the height of the cloud and the starting place for the hailstone*
> T: Good. **You could have also said 'the magnitude of the displacement of the hailstone'**.

Note that we are not yet attempting to distinguish the intentions behind reformulations. Instead we are focusing on when to revoice vs. reformulate. Here, revoicing is similar to the confirmation/clarification questions used in spoken dialogue systems when recognition of the user's speech is uncertain.

We have also implemented a standalone prototype of simple repetitions using the XSLT approach described in (Wilcock, 2001). By this we mean that the system will incorporate part of a student's answer into correctness feedback; e.g. for positive feedback as found in (Chi and Roy, 2010; Dzikovska et al., 2008):

> T: in which direction is the acceleration?
> S: it is [to the right|eastward]
> T: Yes, it is [to the right|eastward]

and, in addition, for negative feedback similar to splicing in (Person et al., 2003); e.g.:

> T: in which direction is the acceleration?
> S: it is [to the right|eastward]
> T: No, if it were [to the right|eastward] then the velocity would be increasing instead of decreasing.

We are currently in the process of completing our analysis of reformulations and are just beginning to implement the approach for revoicing and reformulation we outlined for the Rimac system.

References

I. Beck, M. McKeown, C. Sandora, L. Kucan, and J. Worthy. 1996. Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, 96(4):385–413.

L. Becker, W. Ward, S. Van Vuuren, and M. Palmer. 2011. Discuss: A dialogue move taxonomy layered over semantic representations. In *IWCS 2011: The 9th International Conference on Computational Semantics*, Oxford, England, January.

M. T. H. Chi and M. Roy. 2010. How adaptive is an expert human tutor? In *Intelligent Tutoring Systems Conference, ITS 2010*, pages 401–412.

M. Dzikovska, G. Campbell, C. Callaway, N. Steinhauser, E. Farrow, J. Moore, L. Butler, and C. Matheson. 2008. Diagnosing natural language answers to support adaptive tutoring. In *Proc. of International FLAIRS Conference*.

R. Freedman. 2000. Using a reactive planner as the basis for a dialogue agent. In *Proc. of International FLAIRS Conference*.

P. Jordan, B. Hall, M. Ringenberg, Y. Cui, and C.P. Rosé. 2007. Tools for authoring a dialogue agent that participates in learning studies. In *Proc. of AIED 2007*.

S. Katz, D. Allbritton, and J. Connelly. 2003. 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.

S. Katz, P. Albacete, P. Jordan, and D. Litman. 2011. Dialogue analysis to inform the development of a natural-language tutoring system. In *Proc. of SemDial 2011 (Los Angelogue) Workshop on the Semantics and Pragmatics of Dialogue*.

S. Murillo. 2008. The role of reformulation markers in academic lectures. In A.M. Hornero, M.J. Luzón, and S. Murillo, editors, *Corpus Linguistics: Applications for the Study of English*, pages 353–364. Peter Lang AG.

M.C. O'Connor and S. Michaels. 1993. Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology & Education Quarterly*, 24(4):318–335.

N. Person, A. Graesser, R. Kreuz, and V. Pomeroy. 2003. Simulating human tutor dialog moves in autotutor. *International Journal of Artificial Intelligence in Education*, 12(23-39).

K. VanLehn, C. Lynch, K. Schultz, J. A. Shapiro, R. H. Shelby, and L. Taylor. 2005. The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, 3(15):147–204.

M. A. Walker. 1996. The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence Journal*, 85(1-2):181–243.

G. Wilcock. 2001. Pipelines, templates and transformations: Xml for natural language generation. In *1st NLP and XML Workshop*, page 18.