Extended Explanations as Student Models for Guiding Tutorial Dialogue

Pamela W. Jordan, Maxim Makatchev and Umarani Pappuswamy
Learning Research and Development Center
University of Pittsburgh
Pittsburgh PA 15260
pjordan@pitt.edu

Abstract

The Why2-Atlas tutoring system (VanLehn *et al.* 2002) presents students with qualitative physics questions and encourages them via natural language dialogue to explain their answers. We describe changes we are making to the current version of Why2-Atlas to better utilize a proof-based representation of student essays for motivating dialogue. The abductive proof of the student's explanation for one physics problem serves as a very localized model of likely student beliefs. We will discuss how we can use such a localized student model to select dialogue topics (as the system currently does), to improve the accuracy of the model via clarification requests and to acknowledge and take advantage of the content of the student's essay during the dialogue.

Introduction

The Why2-Atlas tutoring system (VanLehn *et al.* 2002) presents students with qualitative physics problems and encourages them to write their answers along with detailed explanations. Our goal is to have the system engage the student in a meaningful dialogue about the explanation he or she produced. To accomplish this we claim that the system's understanding of the explanation must be represented in such a way that it can not only motivate the dialogue topics the system proposes to the student but also result in a dialogue that references the actual explanation. Although the current version of the Why2-Atlas system is already analyzing student essays and using the results of the analysis to select high-level dialogue goals, the ensuing dialogue is independent of the actual essay.

For example, consider a typical student explanation shown in (1). This essay is taken verbatim from our corpus[1] of human-human computer-mediated tutoring sessions (i.e. it appears exactly as the student entered it) and illustrates the explanations the system is attempting to understand and discuss with the student.

(1) Question: Suppose you are running in a straight line at constant speed. You throw a pumpkin straight up. Where will it land? Explain.

Explanation: Once the pumpkin leaves my hand, the horizontal force that I am exerting on it no longer exists, only a vertical force (caused by my throwing it). As it reaches it's maximum height, gravity (exerted vertically downward) will cause the pumpkin to fall. Since no horizontal force acted on the pumpkin from the time it left my hand, it will fall at the same place where it left my hands.

With this particular essay, the system will identify a high priority goal of resolving the misconception that a force is necessary for the continued motion of an object. Currently the system will attempt to satisfy this goal with a dialogue like that shown below.

(2)

T: OK, let's consider the relationship between force and acceleration in a different context. A hockey puck is resting on the rough ice surface of a skating rink (a hockey team has just finished practicing on it). You push with a force, F, during the time from t1 to t2. Assume that the only force acting horizontally on the puck after the push is friction due to the ice surface (there is no air resistance). What will happen to the velocity of the puck after your push?

S: It slows down over time.

T: Yes. So the puck decelerates. Now assume a Zamboni smooths the ice so that the friction caused by the roughness of the ice surface is reduced. You push the same hockey puck with the same force as before (and for the same amount of time). How does the deceleration of the hockey puck after this push compare to its deceleration following the initial push?

S: It slows down more slowly.

T: Correct. Now assume there exists a machine that can make the ice surface perfectly smooth, so that the friction caused by the ice surface is zero. Again assume the force of air resistance is zero. If you were to push a puck on this surface, what would happen to the velocity of the puck following the push?

S: It doesn't slow down.

T: Right. In general, according to Newton's first law, if there are no forces acting on a body, what happens to the body's velocity?

S: It is constant.

Copyright © 2003, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

[1]The WHY corpus consists of three types of data : (a) Pilot study (b) Spring 2002 evaluation and (c) Fall 2002 pilot. (a) is from uncontrolled experiments, whereas (b) and (c) are from controlled experiments. By controlled we mean that the experiment involved pre and post tests and a control condition. Also, the students who participated in (a) and (b) had completed a prior introductory physics course while those in (c) had never had a physics course.

T: Very good. Something in your essay conflicts with this. Try to revise your essay now, keeping in mind what we just discussed.

An arguably better dialogue would be one that references the student's actual explanation as in the hypothetical dialogue shown below. It is perhaps worth noting that this particular misconception is stubbornly persistent so the student is likely to need help with this across several problems.

(3)

T: Consider the pumpkin's motion before you throw it. What is its horizontal velocity as you are running?

S: constant

T: Right. So what does that tell you about the pumpkin's horizontal acceleration?

S: It is zero.

T: Yes. Now consider Newton's 2nd law in this case. What does that tell you about the horizontal forces on the pumpkin?

S: They sum to zero.

T: Correct. Notice that this conflicts with what you said in your first sentence. Think about this some more and try revising your explanation.

Motivating Tutorial Dialogue

When analyzing student essays, the Why2-Atlas system constructs abductive proofs of the essay because a proof-based approach gives insight into the line of reasoning the student may be following across multiple sentences. This is because proofs of the propositions from the essay share subproofs. Indeed, one proposition's entire proof may be a subproof of the next proposition.

The proofs that Why2-Atlas produces represent the student's knowledge and beliefs about physics with respect to the problem to which he is responding. Acquiring and reasoning about student beliefs and knowledge is one of the central issues addressed by work in student modelling. A student model is a type of user model and in general a user model provides information the system can use in adapting to the needs of its user (Wahlster & Kobsa 1989). In the case of Why2-Atlas, the system needs this representation to identify communicative strategies and goals that will 1) effectively help the student realize and correct his errors and misconceptions and 2) enable the student to realize what reasoning is necessary when generating a complete explanation.

One difficulty any such system must address is uncertainty about the beliefs and knowledge it should attribute to a student. This uncertainty arises because some of the knowledge and beliefs about the student are inferred based on observed student actions or utterances. So as with decision theoretic approaches (Murray & VanLehn 2000; Keeney & Raiffa 1976), the system needs to reason about the utility of separately attributing mutually exclusive representations of varying plausibility to the student. Why2-Atlas tries to estimate this by associating costs with the proofs it creates. However there can still be multiple proofs that are considered equally good representations. We propose that the process of interactively building a proof that represents the student's beliefs and knowledge of physics can also motivate the generation of clarification requests and additional diagnostic questions that explore and confirm the student's line of reasoning.

Once the model is built and uncertainties about it decreased by asking for clarifications from the student, then errors and incompleteness in the reasoning represented can be addressed. A consideration here in identifying and devising strategies to make corrections is that a student's self discovery of errors may be more effective than always being immediately told of the error and its correction. Currently in Why2-Atlas, if the proof reveals a misconception or error then the system will engage the student in a dialogue that works through an analogous, but simplified problem and summarizes at the end with a generalization of the reasoning that the student is expected to transfer to the current problem, as in (2). If incompleteness is revealed by the proof then the system will engage the student in a dialogue that leads the student to express the missing detail. It does so by reminding the student of an appropriate rule of physics, a fact that is relevant to the premise or conclusion of the rule and then asking the results of applying the rule.

Working through an analogous problem is the only technique for leading a student to recognize his error or misconception currently implemented in the system. Another possibility is to step through the student's reasoning as represented by the proof and ask the student to supply inferred details. Having some of these details wrong may have led the student to draw a wrong conclusion and making them explicit may enable her to more easily see the source of her error. As the representation of the explanation was built the system made domain inferences that help to explain the student's statements. In those cases in which there is much evidence for a particular inference the system can commit to accepting that the student has also made this inference but in cases where the system must make many assumptions then it must be skeptical and ask questions that will elicit more details from the student about the inferences that he or she may have made.

Other techniques for dialogue strategies to correct misconceptions, errors and incompleteness relative to proofs may be derivable from argumentation strategies used in argument generation as described in (Zukerman, McConachy, & Korb 2000) (e.g. reductio ad absurdum and premise to goal).

The abductive proof of the student's explanation for one physics problem can serve as a very localized model of likely student beliefs, some of which have not yet been explicitly expressed. Across problems we conjecture that one could merge these snapshots of a student's physics beliefs to create a more general model of the student but for this paper we will focus only on proofs of essays as localized student models. We will discuss how we can use these localized models not only to motivate the dialogue topics the system proposes (as it currently does) but also to improve the accuracy of the model and acknowledge and take advantage of the content of the student's essay during the dialogue.

The cognitive and pedagogical motivations underlying the Why2-Atlas system are further described in (VanLehn *et al.*

2002) and the details of how Why2-Atlas turns students' natural language explanations into proofs are covered in (Jordan & VanLehn 2002).

Background on Building Abductive Proofs

To understand how one might use an abductive proof as a localized student model, it is first necessary to understand at a high-level what weighted abduction is and how abductive proofs are constructed by our system.

Abduction is a process of reasoning from an observation to possible explanations for that observation. In the case of the Why2-Atlas system the observations are what the student said and the possible explanations for why the student said this are the physics qualitative rules (both good and buggy ones) and orderings of those rules that support what the student said. To arrive at the explanation, some assumptions have to be made along the way since all the inferences that underlie an explanation will not be expressed.

Weighted abduction is one of several possible formalisms for realizing abductive reasoning. With weighted abduction there is a cost associated with making an assumption during the inference process. In Why2-Atlas we use Tacitus-lite+ to construct weighted abductive proofs. Following the weighted abductive inference algorithm described in (Stickel 1988), Tacitus-lite+ is a collection of rules where each rule is expressed as a Horn clause. Further, each conjunct p_i has a weight w_i associated with it, as in (4). The weight is used to calculate the cost of assuming p_i instead of proving it where $cost(p_i) = cost(r) * w_i$. The costs of the observations are supplied with the observations as input to the prover.

(4) $p_1{}^{w_1} \wedge \cdots \wedge p_n{}^{w_n} \rightarrow r$

Given an observation or subgoal to be proven, Tacitus-lite takes one of three actions; 1) assumes it at the cost associated with it 2) unifies it with a literal that is either a fact or has already been proven or assumed (in the latter case the cost of the resultant literal is counted once in the total cost of the proof, as the minimum of the two costs) 3) attempts to prove it with a rule.

The applications builder can set cost thresholds and bounds on the depth of rules applied in proving an observation and on the global number of proofs generated during search. Tacitus-lite+ maintains a queue of proof states where the initial proof state reflects that all the observations are assumed. Each of the three above actions adds a new proof state to the queue. The proof building can be stopped at any point and the proofs with the lowest cost can be extracted from the proof state queue. These lowest cost proofs represent the candidates for the most plausible proofs for the observations.

Tacitus-lite+ also includes a fixed set of integrity constraints that are appropriate to our task of modelling the student's reasoning. A distinctive feature of this task is the need to account for erroneous facts and rules. Some false facts correspond to a wrong idealization of the problem and the rest are typically conclusions that students make via the application of false domain rules. Both are modeled by the pairing of buggy domain rules and buggy meta-knowledge rules with their respective correct counterparts. To maintain consistency we implemented a constraint at the meta-level in which the prover prevents the simultaneous appearance of both members of any paired rules in the same proof. Note that while this consistency constraint is natural in theorem proving, from the point of view of student modelling it does represent a risky assumption; that student does not simultaneously hold inconsistent beliefs.

An Example of Alternative Abductive Proofs

Figures 1 and 2 are examples of two simplified alternative abductive proofs for sentence (5). For this sentence, take it as given that the air resistance is 0[2] and assume that it has already been established in another part of the proof that the runner is not applying a horizontal force to the pumpkin after he throws it.

(5) The pumpkin slows down.

Each level of downward arrows from the gloss of a proposition in the two alternative proofs shown in Figures 1 and 2 represent an abductive inference via a domain rule that can be used to prove that proposition or via unification with a right hand side of such a rule (the top level inference in both proofs). To simplify the example we will assume that the weights in all the rules are evenly divided between the propositions in its left-hand side. The number in parentheses at the end of each proposition represents the cost of assuming the proposition.

In the case of both of the proofs shown in Figures 1 and 2, one way to prove that (an axial component or the total) velocity of the pumpkin is decreasing is to unify the goal with the proposition "the horizontal component of the velocity vector is decreasing" so that a rule with the latter proposition in its right hand side becomes applicable. The system will also build alternative proofs in which it will try to interpret the top level goal as the vertical velocity and the total velocity instead (especially since during certain time intervals this is true) but for this example we will ignore these other alternatives for now. Currently both generalizing and specializing unifications are free of charge as shown by the retained cost of 1 of the first subgoal of the proof.

Next we will consider two ways of proving that the horizontal component is decreasing. First let's consider just the case of the proof in Figure 1. In this case Tacitus-lite+ has selected a buggy physics rule that is one manifestation of the impetus misconception; the student thinks that a force is necessary to maintain a constant velocity. In this proof it is assumed that the student has this bug at a cost of .5 and no further attempts are made to prove it. Next Tacitus-lite+ proves that the total force on the pumpkin is zero by proving that the possible addend forces are zero. Since it is a given that air resistance is negligible this proposition unifies with this given fact for zero cost. Likewise, since we said that it was already proven elsewhere that the man is applying a horizontal force of 0 to the pumpkin after he throws it, this proposition unifies with the proven fact for zero cost as

[2] Students often overlook relevant givens, so proofs that ignore a given can be generated as well.

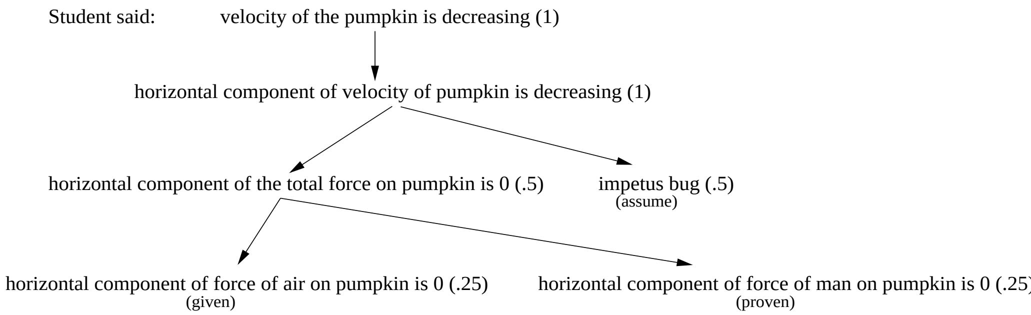

Figure 1: Example of One Possible Simplified Abductive Proof for "The pumpkin slows down."

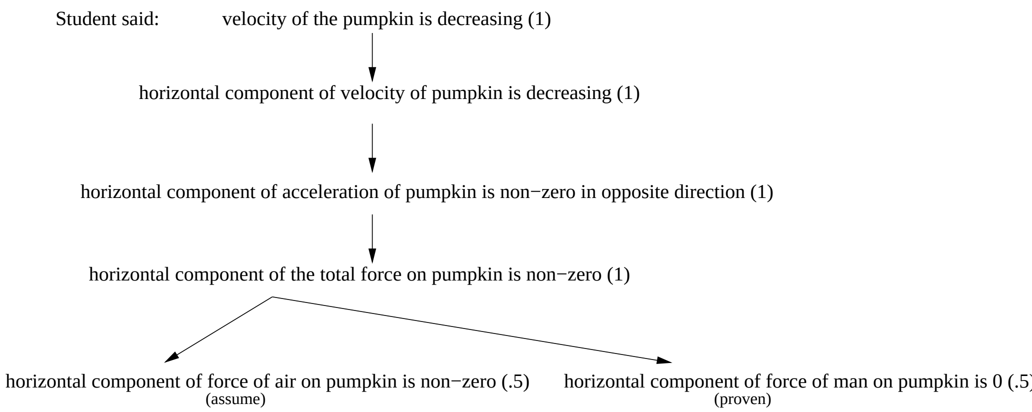

Figure 2: Example of an Alternative Simplified Abductive Proof for "The pumpkin slows down."

well. Since the proof contains just one assumption, that the student has the impetus bug, the total cost of the proof is .5.

Looking again at the alternative proof in Figure 2, we see that it attempts to prove the horizontal component of the velocity is decreasing by first trying to prove that the horizontal component of the acceleration is non-zero and to prove this it is necessary to prove that the total horizontal force on the pumpkin is non-zero. One way to prove this is to prove that exactly one of the addend forces is non-zero. The system can ignore the given at this point in order to try to prove that there is exactly one non-zero addend force on the pumpkin. Namely it tries to prove that wind resistance is not negligible but since it cannot prove this it must be assumed at a cost of .5. So the total cost of this alternative proof is .5 as well. In this example, the system now has two plausible proofs with no means of choosing between them without more information from the student.

Current System Behavior

Currently the system arbitrarily picks one of the best proofs in order to select high-level dialogue goals. Once a single proof is selected, the system applies a set of test patterns to parts of the proof to identify possible dialogue goals. It can test for combinations of patterns for givens (mainly to get bindings for variables in a pattern), for assumed propositions, for propositions asserted in the student's essay, and for inferred propositions. In addition it can also test for missing patterns in the proof and for particular domain rules to have been used. Each dialogue goal that the system is capable of addressing is linked to sets of patterns that are expected to be indicative of it. In the case of the proof in Figure 1 it identifies a dialogue goal to address the impetus misconception since an impetus bug assumption is part of the proof. In the case of the proof in Figure 2 it identifies a goal to address the wrong assumption that air resistance is non-negligible.

In addition to selecting high-level dialogue goals, the system can also already give some direct, constructive feedback on an essay. It does so via test patterns that target important details that students often leave out. For example, with "the pumpkin slows down", it will ask the student to clarify whether she means the horizontal or vertical motion. However, our expectation is that writing specific tests for missing details will become unnecessary once we extend the system to treat abductive proofs as localized student models.

A Student Model Treatment of Abductive Proofs

There are two situations that we need to address to better utilize the abductive proofs generated. The first is when the system generates multiple lowest cost alternative proofs as in the above example associated with Figures 1 and 2. The second is when it produces just one lowest cost proof in which the student has presented either unambiguous correct information in the essay that leads the system to infer a correct and complete explanation or the student has provided information that is clearly incorrect.

In the situation where there are multiple lowest cost proofs, the goal is to reduce some of the uncertainties about what the student's beliefs and knowledge are. To do this, the system needs to locate major differences between the alternative proofs and ask clarification questions that will help the system determine which proof is the best representation of the student's beliefs. For example in the above example with Figures 1 and 2, there are a number of possibilities for doing this. First the system could simply ask "Why do you think the pumpkin slows down?". If the student responds with "I don't know." then the system can ask a more leading question such as "What is the total horizontal force on the pumpkin?" since a proposition about total force is present in both proofs but has different values for the variables.

The other situation is where the system finds one plausible proof either on its own or because it asked questions that allowed it to narrow down the possibilities to one proof as above. When the system has to infer most of the proof from just a few true student propositions, one possible strategy is to simply move from an explicit student proposition to some proposition that is at a distance N from it on a path in the proof tree. For example, if the student had said "The pumpkin lands on me because the velocity of the pumpkin is constant." and N is 1, the system might say "Please explain more about why this is so. What does the constant velocity tell us about the acceleration of the pumpkin?". Another strategy might be to find a proposition that is midway between two explicit propositions and try to elicit that proposition.

If on the other hand the one plausible proof contains a bug, because the student made some incorrect statements as in the essay in (1), the system could treat these incorrect student statements as being correct and attempt to lead her to a contradiction (e.g. reductio ad absurdum). For example, the target starting point might be "There was a horizontal force acting on the pumpkin before the throw" with it leading to the statement "the horizontal velocity of the pumpkin is not constant before the throw" which contradicts the conjunction from the problem statement that "the man is running with a constant speed" and "the man is carrying the pumpkin." Another alternative might be to simply elicit correct propositions from the student by starting from any true propositions that student may have included in the essay or from givens as with the dialogue example in (3).

Work on natural language interpretation selection (Zukerman & George 2002) and natural language generation work on argumentation and explanation (e.g. (Zukerman, McConachy, & Korb 2000; Horacek 1997; Zukerman & McConachy 1993)) will give us a starting point for deriving feedback to the student relative to a proof but there are a number of new challenges to address. First, generation work on argumentation and explanation has focused on generating text as opposed to dialogue and deals mainly with correct representations. And in the case of choosing between alternative representations, the work in (Zukerman & George 2002) doesn't address the problem of choosing between representations that have the same number of inferences as is the case for the alternative proofs in Figures 1 and 2.

Generating the Tutor's Contribution

The appropriateness of the dialogue strategies described above will depend on the structure of the proofs generated. But since this may not be enough to uniquely identify one best dialogue strategy, we will randomly pick from the set of candidates in our initial implementation. However, to create a more informed selection of strategies we will also consider using additional features of the dialogue and the student such as has been done with fuzzy selection rules in (Person *et al.* 2001) and reasoning about the utilities of strategies as is done with tutorial actions in (Murray & VanLehn 2000).

Once a strategy is selected the system must produce surface realizations that will satisfy the goals of the strategy. Surface realization in the current version of Why2-Atlas is accomplished via either strings that are attached to goals or template-based generation. Template generation will be our preferred method for realizing the dialogue strategies discussed above. Our current implementation of template generation takes as input a communicative goal and the physics concepts involved and uses an appropriate template to construct the deep syntactic structures required by the RealPro realizer (Lavoie & Rambow 1997). (Wilson 2003) addresses the problem of automating the process of template building for questions.

Current Status of New Work

We are currently reviewing proofs of the student essays in our corpus in order to refine the above suggested strategies. So far we have confirmed that the *reductio ad absurdum* strategy, in which the student is lead to a contradiction, is used frequently in human-human dialogues. One such dialogue, which is taken from our corpus of human-human computer-mediated tutoring sessions, is shown in (6).

(6)

[:C1] S: If I were to throw the pumpkin straight up, while running with a constant speed, the pumpkin would land behind me.
[:C1] contradicts [:C2]

T: Can you explain why?

S: While I am running, both the pumpkin and I are moving with the same speed. Once I throw the pumpkin up, it no longer has anything to thrust it forward in the either the horizontal or vertical direction. Therefore, it will fall to the ground behind me.

T: Does an object always need something to thrust it to be moving?

S: An object always needs a thrust or push to keep it accelerating but from this problem, we know that it has constant speed which means no acceleration.

T: Yes, the object needs a thrust to accelerate. But does it need a thrust to be moving at constant speed?

S: No, Newton's first law tells us that an object in motion will continue in motion until an outside force acts upon it. Simarily, an object in rest will continue in rest unless an outside force acts upon it.

T: Tutor: Good. Now consider the pumpkin just after throwing it up. What forces are acting on it?

S: After the pumpkin is thrown up, gravity is working on the pumpkin.

T: What direction does it act?

S: Gravity acts in the vertical direction.

T: Neglecting friction then there is only a vertical force acting on the pumpkin, right?

S: Yes.

T: Now tell me whether this will affect the horizontal motion of the pumpkin

S: This will not affect the horizontal motion of the pumpkin because gravity only acts in the vertical direction.

T: Right. What was the horizontal velocity of the pumpkin just before it was thrown up?

S: Velocity is a vector so it has both a magnitude and a direction. The pumpkin velocity just before it was thrown up was constant with my velocity.

T: Right again. Since you and the pumpkin have the same horizontal velocity, which is constant, where do you think the pumpkin will land?

S: Is this like the example where if you shoot a bullet from a moving convertible, the bullet will land back in the gun?

T: Absolutely

[:C2] S: Ok, if I were to thrown the pumpkin straight up while I was running with a constant speed, the pumpkin would land on me– ouch!

T: Sorry about it, but now you revise the essay

A detailed examination of 18 dialogues from our corpus revealed that there were 17 contradictory statements made by the students in their discussions with 3 tutors. 7 of these cases were related to contradictions of the student's initial answer and 10 were related to contradictions of physics concepts held by the student. The 3 tutors represented in the set of dialogues successfully helped the students resolve wrong assumptions, misconceptions and errors by a strategy of leading them to a contradiction. Once we have identified and confirmed additional strategies that would appear to work for a majority of the essays in our corpus, we will begin to implement and test these strategies in the Why2-Atlas system.

Acknowledgments

This research was supported by MURI grant N00014-00-1-0600 from ONR Cognitive Science and by NSF grant 9720359. We thank the entire NLT team for their many contributions in creating and building the Why2-Atlas system. In particular we thank Michael Ringenberg for his work with Tacitus-lite+, Michael Böttner for his ontology work and Roy Wilson for his work in sentence realization.

References

Horacek, H. 1997. A model for adapting explanations to user's likely inferences. *User Modeling and User-Adapted Interaction* 7(1):1–55.

Jordan, P. W., and VanLehn, K. 2002. Discourse processing for explanatory essays in tutorial applications. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*.

Keeney, R., and Raiffa, H. 1976. *Decisions with multiple objectives*. Wiley.

Lavoie, B., and Rambow, O. 1997. A fast and portable realizer for text generation systems. In *Proceedings of the Fifth Conference on Applied Natural Language Processing Chapter of the Association for Computational Linguistics*, 265–268.

Murray, R. C., and VanLehn, K. 2000. DT tutor: A dynamic decision-theoretic approach for optimal selection of tutorial actions. In *Intelligent Tutoring Systems, 5th International Conference*.

Person, N. K.; Graesser, A. C.; Kreuz, R. J.; Pomeroy, V.; and the Tutoring Research Group. 2001. Simulating human tutor dialog moves in AutoTutor. *International Journal of Artificial Intelligence in Education* 23–29.

Stickel, M. 1988. A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Technical Report 451, SRI International, 333 Ravenswood Ave., Menlo Park, California.

VanLehn, K.; Jordan, P. W.; Rosé, C.; Bhembe, D.; Böttner, M.; Gaydos, A.; Makatchev, M.; Pappuswamy, U.; Ringenberg, M.; Roque, A.; Siler, S.; Srivastava, R.; and Wilson, R. 2002. The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In *Proceedings of the Intelligent Tutoring Systems Conference*.

Wahlster, W., and Kobsa, A. 1989. User models in dialogue systems. In Kobsa, A., and Wahlster, W., eds., *User Models in Dialogue Systems*. Springer Verlag, Berlin. 4–34.

Wilson, R. 2003. A method for comparing fluency measures and its application to its natural language question generation. In *Proceedings of AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*.

Zukerman, I., and George, S. 2002. A minimum message length approach for argument interpretation. In *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*.

Zukerman, I., and McConachy, R. 1993. Generating concise discourse that addresses a user's inferences. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1202–1207. Morgan Kaufmann Publishers, Inc.

Zukerman, I.; McConachy, R.; and Korb, K. B. 2000. Using argumentation strategies in automated argument generation. In *Proceedings of the 1st International Natural Language Generation Conference*, 55–62.