# Good computational manners: Mixed-initiative dialog in conversational agents

**Max Louwerse, Art Graesser, Andrew Olney**

**and the Tutoring Research Group**

Department of Psychology, Institute for Intelligent Systems
University of Memphis
Memphis, TN 38152
{mlouwers, a-graesser, aolney} @memphis.edu

**Abstract**

Polite behavior in intelligent conversational systems requires mixed-initiative dialog. AutoTutor is an intelligent tutoring system in which several of these conversational skills are implemented. This paper discusses some aspects of mixed-initiative dialog, as implemented in AutoTutor ranging from discourse markers, dialog moves and speech acts to question answering. In addition some of the studies will be reported that show the effectiveness of AutoTutor's conversational skills on pedagogy, conversational smoothness and learning gains.

## Introduction

In human-computer interaction computers help humans in the execution of intelligent actions. One of these intelligent actions humans are good at is communication. Humans are for instance excellent conversationalists. We chat, talk, gossip, discuss. In fact, the way we learn is often by some form of communication. In those cases where natural language is involved, it therefore seems sensible to develop intelligent systems that allow for human-computer interaction that is as natural as possible to human-human communication. Mixed-initiative dialog should be part of such a conversational system.

## Mixed Initiative Dialog

Participants in conversations are expected to cooperate; in other words, they are expected to adhere to the cooperative principle (Clark, 1996; Grice, 1975). Their contributions in the conversation are supposed to be such as is required, at the time of the conversation by the purpose or direction of that conversation. For computers, more specifically intelligent tutoring systems, participating in such a joint activity is difficult. At the same time, however, to express acceptable behavior in a conversation, mixed-initiative engagement in that conversation is essential.

Mixed-initiative interaction refers to the flexibility of strategies applied in the interaction between participants by those participants. Mixed-initiative interaction allows for the direction and control of the interaction to be shifted between participants. At least four levels of mixed-initiative interaction can be distinguished (Allen, 1999).

1. *Unsolicited reporting*, in which a participant notices problems during the interaction and notifies the other participant of critical information;
2. *Subdialog initiation*, in which the participant initiates a subdialog to clarify and correct information;
3. *A fixed-subtask mixed initiative*, in which the participant takes the responsibility to perform certain operations like solving predefined subtasks.
4. *Negotiated mixed-initiative dialog*, in which participants coordinate and negotiate to determine initiative.

Mixed-initiative dialog requires that the participants engaged in the joint project (i.e. computer and human) act in coordination with each other. For human-initiated interactions (like booking a flight on the web) or system-initiated interactions (like checking your bank balance using the telephone keypad) this is relatively easy. One participant, either the user or the system, always has control in the interaction (production and comprehension) of information and directs the discourse. These are examples of Allen's first two levels of interaction. Reaching the third and fourth level of mixed-initiative interaction is a real challenge for intelligent system.

True mixed-initiative dialog requires active participation in the joint project from both participants, which involves more than an understanding of the linguistic (syntactic and semantic) information of the input is needed. It requires knowledge of its pragmatics. Strictly speaking, although intelligent systems have reached a level sophistication in analyzing locutionary acts (using syntactic parsers and information retrieval procedures), they are far from understanding illocutionary acts, and accordingly performing perlocutionary effects. However, as we know from human communication, all three (locutionary, illocutionary and perlocutionary aspects) are important for mixed-initiative dialog (see Clark, 1996).

Perhaps the question to be addressed in mixed-initiative interaction in intelligent systems is not whether models of human-computer interaction *are* mixed-initiative, but how the *impression* can be created that users are involved in mixed-initiative dialog. We claim that the impression of mixed-initiative can be created by making the computer user believe that they are participating in a natural conversation. The more natural this conversation seems to be, the more likely the impression of mixed-initiative dialog is formed.

In the development of an intelligent tutoring system we considered mixed-initiative dialog, or the impression thereof, as a prerequisite for a conversational tutor (see Appendix A). The remainder of this paper presents various aspects of mixed-initiative dialog as they are currently implemented in the intelligent tutoring system AutoTutor, and discusses their effects on the conversation and learning.

## AutoTutor

AutoTutor is a conversational agent that assists students in actively constructing knowledge by holding a conversation in natural language. At least four components can be distinguished in the system (Graesser et al. 1999).

1. AutoTutor uses Latent Semantic Analysis for its world knowledge. LSA uses singular value decomposition to reduce a co-occurrence matrix of words (or documents) to a cosine between two vectors. In particular, AutoTutor uses LSA to give meaning to a student answer and to match that answer to ideal good and bad answers (see Franceschetti et al., 2001).
2. A dialog management system guides the student through the student computer exchange accommodating student input. Fuzzy production rules and a Dialog Advancer Network form the basis of these conversational strategies (see Person et al, 1999).
3. For its didactic skills AutoTutor uses curriculum scripts that organize the pedagogical macrostructure of the tutorial. These scripts keep track of the topic coverage and follow up on any problems the student might have (see Graesser et al, 2002).
4. A talking head with facial expressions and synthesized speech is used for the interface. Parameters of the facial expressions are generated by fuzzy production rules (see Graesseret al., 2002).

An overview of the interface that shows the talking head, the question box, the answer box and the dialog history, is presented in Appendix B.

One of the advantages of AutoTutor is that its conversational skills are independent of the domain it tutors in. The system can thus be seen as a collaborative scaffold. In fact, AutoTutor was designed to be reusable for other knowledge domains that do not require mathematical precision and formal specification. This is an important aspect of the system, since its conversational components remain intact when domains are changed.

In order to test the portability of the AutoTutor architecture, we developed a version for the domain of conceptual physics. Together with computer literacy, conceptual physics is one of the fields in which extra tutoring sessions are needed. The target population for the tutor was undergraduate students taking elementary courses in conceptual physics. In the transition of AutoTutor from computer literacy to physics only three modules needed to be changed for the new subject matter: (1) a glossary of terms and definitions for physics, (2) an LSA space for conceptual physics, (3) a curriculum script with deep reasoning questions and associated answers for physics. The three modules can loosely be affiliated with metacognition, comprehension, and production. Note that all mixed-initiative dialog components remained intact.

We will discuss four of AutoTutor's components that form the structure of mixed-initiative interaction: 1) discourse markers, 2) dialog moves, 3) speech acts, and 4) question answering.

### Discourse markers

Discourse markers are cues that facilitate the transition between turns in a dialog. They support the conversational smoothness of the conversation. For intelligent conversational systems the main problem lies in choosing the right discourse marker at the right time in order to give the impression of conversational smoothness.

To start building a data-driven taxonomy of cohesion relations in dialog, we used the methodology described in Knott & Mellish (1996). Knott argues for a data-driven account of taxonomies by proposing a substitution test, to see whether one cue phrase (a relational phrase cueing a coherence relation) can be used in the place of another. Two cue phrases are considered more or less synonymous within the same category if they are inter-substitutable. If one can be replaced by the other, but not the other way around, the latter is a hypernym of the former. If they cannot be substituted in any given context they are exclusive. After making a large number of substitutions that were then entered into a factor analytic model, four categories emerged: direction, polarity, acceptance and empathy (see Louwerse & Mitchell, under review). The taxonomy that could be constructed out of these markers was implemented in AutoTutor. Random selections from large bags of discourse markers were made in order to give positive, neutral and negative feedback to the student, and to provide the student with a conversational continuity that can be found in natural dialog. These selections are based on the Dialog Advancer Network and LSA scores. For instance, a high LSA score can lead to positive feedback like 'Excellent! Well done!', to neutral feedback like 'Hmm, okay' or to negative feedback like 'No, not really'. Using the taxonomies of markers allowed for the impression of conversational smoothness in AutoTutor. By going beyond simple yes-no feedback the impression was created that the system was not fully in control of the conversation, but allowed for a conversation with the student.

## Dialog moves

A large number of AutoTutor's dialog moves consist of a lengthy substantive contribution that prompts the learner for more information, that adds information, or that corrects a student error. These dialog moves don't particularly make AutoTutor a conversational and mixed-initiative system. More specific dialog moves that mimic natural tutoring sessions however do. AutoTutor primarily uses seven kinds of dialog moves in its conversation with the student: it *prompts* for specific information; it *pumps* for more information; it *hints*, a*sserts, corrects, repeats* and s*ummarizes.* Discourse patterns organize these dialog moves in terms of their progressive specificity. Hints are less specific than Prompts, and Prompts are less specific than Elaborations. AutoTutor therefore cycles through a Hint-Prompt-Elaboration pattern until the student articulates the correct answer (or rather reaches threshold *t* for covering all aspects of the ideal answer). The other dialog moves (e.g., short feedbacks and summaries) are controlled by fuzzy production rules. With the Dialog Advancer Network in which these and other dialog moves are implemented, AutoTutor monitors student input and student progress and steps in when critical information is needed, thus reaching as least the unsolicited reporting mixed-initiative level.

Graesseret al. (1999) conducted a study that evaluated the dialog moves generated by AutoTutor. Student answers (generated by human students) in a computer literacy class were collected. AutoTutor's dialog moves in response to these answers were recorded and these fragments were rated on smoothness and pedagogy by expert raters. Overall, raters concluded that AutoTutor performed well in generating moves that fit in with the flow of the conversation and are conversationally sound. AutoTutor

## Speech act classifier

In natural conversation an understanding of the intentions of the participants is needed in addition to understanding the meaning and syntactic structure of an utterance. Computational intention recognition, however, is difficult. First of all, there are various distinct ways of formulating an intention. Secondly, intentions often remain linguistically unmarked. Finally, classifying a speaker's intention seems to require some underlying framework and existing speech act classifications are very different from one another and are based on different approaches.

Understanding the intentions of participants requires a thorough understanding of the context in which the utterance is expressed, not one of the strong points of intelligent systems. Nevertheless, we have initial evidence that if a speech act is taken out of context, computational speech act classification might be on par, or might be even better than speech act classification by humans. Settling for a system that satisfices, we have developed a speech act classifier that identifies a large number of categories using surface linguistic features. Furthermore, the classifier does not only focus on the uptake of the speech act, but anticipates appropriate response.

The classifier identifies 20 illocutionary categories, ranging from metacommunicative and metacognitive expressions like "Can you repeat that?" and "I don't know" to 17 questions categories, as proposed by Graesser et al. (1992). These categories consisted for instance of definitional questions, causal antecedent questions and example questions. Extensive testing of the classifier showed that the accuracy of the classifier was better than 65% overall and as high as 87% for certain question categories (see Louwerse & Olney, in preparation).

With the classifier successfully been implemented in AutoTutor, the system now has ways to appropriately respond to a student's input. Using the illocutionary acts of the students, it can notify them of critical information (and follow the dialog moves described before) or it can initiate new subdialogs to clarify and correct the student. Recognizing the student's intentions brings AutoTutor closer to simulating mixed-initiative.

## Question answering tool

Illocutionary acts require anticipating perlocutionary effects. AutoTutor appropriately responds to nonsense input, to metacommunicative and metacognitive input, as well as to student contributions. Appropriately responding to student questions, one of the prerequisites for a fully mixed-initiative dialog systems, is however difficult.

Various question answering systems have been developed, most notably systems that compete in the TREC competitions and those developed in the AQUAINT program. Often these systems provide short answers to who-and-what questions. Information needs to be concise and correct. In a tutoring environment deep reasoning questions might prevail. Instead of who-and-what questions, the system will have to be able to answer why questions. Furthermore, information should not be concise. Instead the student needs an elaborative answer that covers various aspects of the answer.

Recently, we have developed a Question Answering Tool (QUANTUM) that answers any student question in the desired format (see Appendix A). The tool combines the surface cue based categorization in the speech act classifier with world knowledge using LSA. By combining these two approaches, the tool selects a paragraph from a document or series of documents as the answer to that question (e.g. Hewitt, 1998). Current performance for relevance and informativity of the answers provided by our system is satisfactory. Experiment using AutoTutor with QUANTUM implemented in the system showed that subjects were satisfied with the answers to their questions.

With the current mechanism we are easily able to answer questions in various domains. Also, those questions that are generally considered difficult to answer computationally (e.g. causal antecedent, comparison questions instead of definitional questions that can answered from glossaries) have the highest performance scores, due to the ideal computational combination of (a) syntactic, lexical, and surface cue features and (b) world knowledge.

## Learning gains in AutoTutor

So far we have discussed various conversational aspects of AutoTutor that support a mixed-initiative dialog. The remaining question of course concerns the performance of the system. Person et al., 2001) tested the learning gains of students who had conversational interactions with AutoTutor. Sixty students in a computer literacy course at the University of Memphis were participated in experiments in one of three experimental conditions: AutoTutor (student interacted with AutoTutor to learn about one of the three computer literacy topics, Hardware, Operating systems, or Internet), Reread (student reread material in the course textbook about one of the three topics), and no-read Control (student does *not* re-read or interact with AutoTutor for one of the three topics). Students were given approximately 50 minutes to reread the material or to interact with AutoTutor.

Person et al. showed that there were significant differences among the three experimental conditions, with means of .43, .37, and .35 in the AutoTutor, Reread, and Control conditions, respectively. Planned comparisons showed the following pattern: AutoTutor > Reread = Control. The effect size of AutoTutor over Control was .50 standard deviations. Given that students spent the same amount of time in the AutoTutor (50.6 minutes) and Reread (49 minutes) conditions this result is encouraging.

Currently additional experiments are conducted with versions of AutoTutor that use more mixed-initiative dialog components that were not available in earlier versions. Furthermore, these experiments will be able to show whether there are differences between domain (computer literacy versus physics).

## Conclusion

Intelligent tutoring systems like AutoTutor that interact with humans are not mixed-initiative. However, by using pedagogical dialog moves to macrostructure the tutorial, by using speech acts to anticipate student’s illocutionary acts and accordingly decide on perlocutionary effects and by using a sophisticated question answering tool, the impression can be given of a mixed-initiative conversational partner. Following Allen’s (1999) mixed-initiative levels, unsolicited reporting, subdialog initiation and fixed subtask initiative are to a certain extent represented into AutoTutor. Although full mixed-initiative dialog might be far away, AutoTutor provides more natural conversations that contribute to the enhanced impression of mixed-initiative dialog. Not only are its dialog moves indistinguishable from a human tutor, its learning gains are similar to those of a human tutor.

In sum, the etiquette for human-computer interaction should follow the skills that humans use continuously, that of conversations. AutoTutor is an example of what such etiquette system could look like.

## Acknowledgements

This research was supported by grants from DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (N00014-00-1-0600) and the National Science Foundation (SBR 9720314 and SBR 0106965).

## References

Allen, J. (1999). Mixed-initiative interaction. IEEE Intelligent Systems. Vol. 14(5), pp.14-16.

Clark, H.H. (1996). *Using Language*, Cambridge: Cambridge University Press.

Franceschetti, D.R., Karnavat, A., Marineau, J., McCallie, G. L., Olde, B. A., Terry, B. L., & Graesser, A. C. (2001). Development of physics text corpora for latent semantic analysis. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 297-300). Mahwah, NJ: Erlbaum

Graesser, A. C., Person, N., & Huber, J. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), Questions and information systems. Hillsdale, NJ: Erlbaum.

Graesser, A.C., Person, N., Harter, D., & TRG (1999). Teaching tactics and dialog in AutoTutor. International Journal of Artificial Intelligence in Education.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2002). Intelligent tutoring systems with conversational dialogue. *AI Magazine*.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group. (1999).

AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*,**1**, 35-51.
Grice, H. P., (1975). Logic and Conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics*, vol. 3, Academic Press.
Hewitt, P. (1998). *Conceptual physics*. Reading, etc., Addison & Wesley.
Knott, A. & Mellish , C. (1996). A feature-based account of the relations signalled by sentence and clause connectives. *Language and Speech* 39, 143-183.
Louwerse, M. & Mitchell, H. (under review), *Toward a Data-driven Taxonomy of Discourse Markers in Dialog.*
Louwerse, M. & Olney, A.. (in preparation). *Toward an automated speech act classifier: Surface cues for categorizing conceptual questions*.
Person, N.K., Graesser, A.C., Kreuz, R.J., Pomeroy, V., & TRG (1999). Simulating human tutor dialog moves in AutoTutor. *International Journal of Artificial Intelligence in Education*.
Person, N.K., Graesser, A.C., Bautista, L., Mathews, E., & TRG (2001). Evaluating student learning gains in two versions of AutoTutor. In J.D. Moore, C.L. Redfield, and W.L. Johnson (Eds.) *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future* (pp. 286-293). Amsterdam: OIS Press.

## Appendix A

**Example of student interaction with AutoTutor**

**Tutor:** The sun exerts a gravitational force on the earth as the earth moves in its orbit around the sun. Does the earth pull equally on the sun? Explain why.

**Student:** How can gravity pull on both the sun and the earth at the same time?

**Tutor:** Very interesting point. The following will probably help. The sun also contributes to ocean tides, although it is less than half as effective as the moon in raising tides - even though its pull on the earth is 180 times greater than the pull of the moon. Why doesn't the sun cause tides 180 times greater than lunar tides? The answer has to do with a key word: difference. Because of the great distance of the sun, the difference in its gravitational pull on opposite sides of the earth is very small. The percentage difference in the sun's pulls across the earth is only about .017 percent, compared to 6.7 percent across the earth by the moon. It is only because the pull of the sun is 180 times stronger than the moon's that the sun tides are almost half as high (180 X 0.017 percent = 3 percent, nearly half of 6.7 percent). Do you understand the answer?

**Student:** no

**Tutor:** What is it that you don't understand?

**Student:** I don't know what gravity is.

**Tutor:** gravity is the attraction between objects due to mass.

**Tutor:** Okay, see if you can get this. Newton's Third Law refers to the forces exerted by one body on another __________?

**Student:** What is Newton's third law

**Tutor:** Wait a minute, answer my question and then we'll do yours. Let me repeat. Newton's Third Law refers to the forces exerted by one body on another __________?

**Student:** body

## Appendix B

### AutoTutor interface

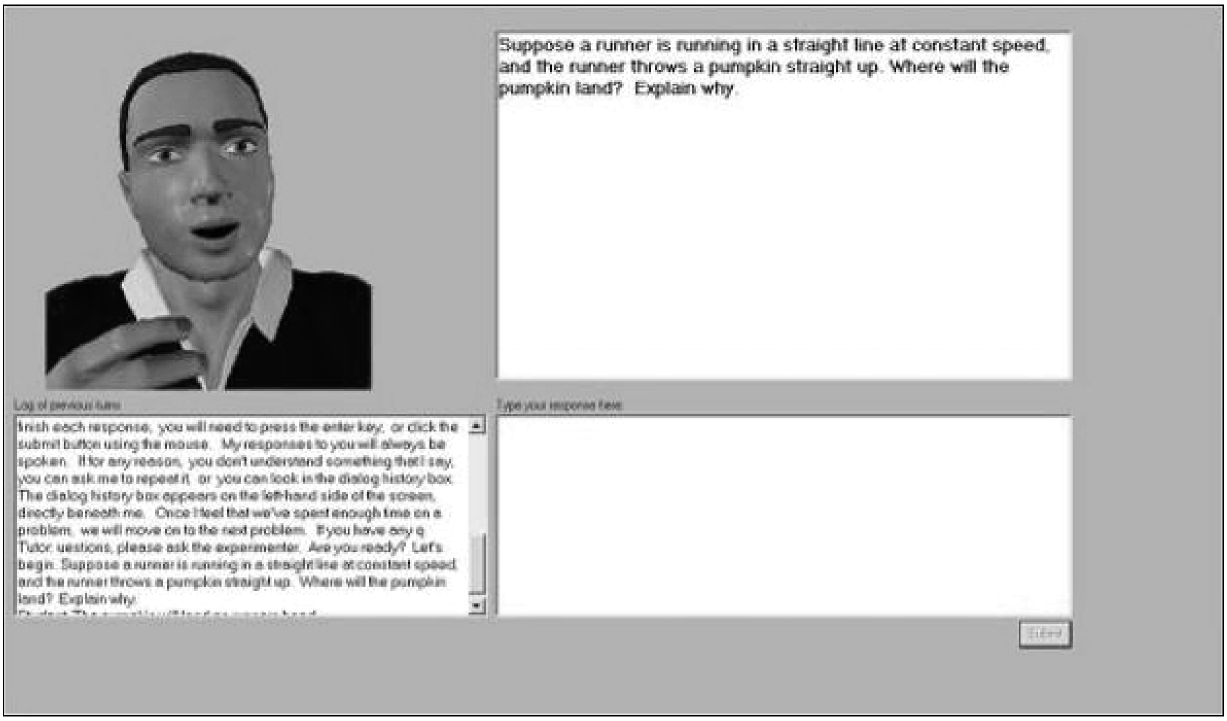

**Figure 1. Overview of AutoTutor interface**