

The Role of Initiative in Tutorial Dialogue

Mark G. Core and Johanna D. Moore and Claus Zinn

School of Informatics

University of Edinburgh, 2 Buccleuch Place

Edinburgh EH8 9LW, UK

[markc|jmoore|zinn]@inf.ed.ac.uk

Abstract

This work is the first systematic investigation of initiative in human-human tutorial dialogue. We studied initiative management in two dialogue strategies: didactic tutoring and Socratic tutoring. We hypothesized that didactic tutoring would be mostly tutor-initiative while Socratic tutoring would be mixed-initiative, and that more student initiative would lead to more learning (*i.e.*, task success for the tutor). Surprisingly, students had initiative more of the time in the didactic dialogues (21% of the turns) than in the Socratic dialogues (10% of the turns), and there was no direct relationship between student initiative and learning. However, Socratic dialogues were more interactive than didactic dialogues as measured by percentage of tutor utterances that were questions and percentage of words in the dialogue uttered by the student, and interactivity had a positive correlation with learning.

1 Introduction

Tutorial dialogue systems face the unique problem that users (students) often do not know the answers to questions asked by the system and may produce wrong answers that are not in the system's domain model. Because of these difficulties, current tutorial dialogue systems are largely system-initiative; only the system asks questions, and for each question, system designers build a database of poten-

tial correct and incorrect answers, and a set of responses to deal with the incorrect answers.

There has been a similar trend in the spoken dialogue systems community. The problem in this case is poor speech recognition performance and the solution is for the system to ask questions with a limited set of answers. However, Chu-Carroll and Nickerson (2000) showed that a suitably intelligent mixed-initiative dialogue system (MIMIC) outperformed a comparable system-initiative dialogue system in terms of user satisfaction and task efficiency. MIMIC could back off to system-initiative mode when necessary but otherwise operate in mixed-initiative mode.

The cognitive science literature indicates that such a breakthrough is also needed in the tutoring community. The current system-initiative approaches conflict with arguments that it is the highly collaborative nature of human-human tutoring dialogue that leads to learning (Merrill et al., 1992a; Fox, 1993; Graesser et al., 1995). Through this dialogue, tutors can intervene to ensure that errors are detected and repaired and that students can work around impasses (Merrill et al., 1992b). Previous research has also shown that students must be allowed to construct knowledge themselves to learn most effectively (Chi et al., 1989; Chi et al., 1994; VanLehn et al., 1998). The consensus from these studies is that experienced tutors maintain a delicate balance allowing students to do as much of the work as possible and to maintain a feeling of control, while providing students with enough guidance to keep them from becoming too frustrated or confused. We refer to this style of tutoring as "Socratic" because it is characterized by the use of questions and other hints to draw out answers from students having difficulty.

(Rosé et al., 2000) gives an overview of the evidence in favor of Socratic tutoring as well as describing an opposing viewpoint supporting a tutoring style referred to as didactic. Here, rather than drawing out the answer from the student, the tutor points out the student's error and explains how to derive the correct answer.

We hypothesized that (1) didactic tutoring corresponds to the system-initiative dialogue management currently implemented in tutorial dialogue systems, (2) Socratic tutoring is mixed-initiative, and (3) furthermore that initiative is directly related to "Socraticness" — more student initiative would mean more student learning although a minimum amount of tutor initiative is likely to be necessary.

To investigate these hypotheses, we undertook a systematic investigation of initiative, tutoring strategy (Socratic vs. didactic), and learning (task success) using a series of human-human tutoring dialogues from an earlier project (Rosé et al., 2000). In one set of dialogues, the tutor used a Socratic tutoring style while in the other she used a didactic tutoring style. We annotated these dialogues for initiative, measured the distribution of initiative in the Socratic and didactic dialogues, and measured the relationship between initiative and student learning.

2 Previous Work

2.1 Defining Initiative

Shah (1997) defines student initiative as "any contribution by the student that attempts to change the course of the [tutoring] session" (p. 13). Shah's corpus analysis dealt with remediation dialogues where a tutor quizzed students about the answers they gave during problem solving. In this corpus, student initiatives are student utterances that are not answers to questions. Shah assumes that these initiatives are dealt with exclusively by the tutor's next speech act, and that initiative then reverts back to the tutor. This definition was too limited for our more free-form tutoring dialogues.

Sinclair and Coulthard (1975) developed a dialogue grammar for classroom discussions. Their minimal unit of dialogue is the *exchange* which is composed of an initiating move, an optional re-

sponding move, and an optional feedback move. Whoever makes the initiating move is said to have initiative for the exchange. Although questions can be reasked in cases of incorrect student answers, this framework does not capture other ways an exchange can be disrupted (*e.g.*, the student asks a question rather than answering the current question), and again this definition was too limited for our dialogues.

Linell et al. (1988) discuss how a responder can ask for clarification, challenge the speaker, and change topics as well as respond directly to an initiating move. Linell et al. do not assign initiative directly to speakers but instead rank speaker moves based on how much "they can be regarded as governing or steering the ensuing dialogue and as being governed or commanded by the preceding dialogue" (p. 419). For example, an utterance which is not a response in any way but requires a response from the listener is ranked highest with a value of six. Minimal responses are at the other end of the scale (with a rank of two); they invite no response and give no more information than required.

Linell et al.'s approach was to sample a wide variety of dialogue genres in developing their definition; in contrast, Chu-Carroll and Brown (1998) focussed specifically on problem-solving dialogues. They found that it was important to differentiate initiative (which they call dialogue initiative) from task initiative. They define dialogue initiative by stating that it "tracks the lead in determining the current discourse focus" (p. 6),¹ and that task initiative "tracks the lead in the development of the agents' plan" (p. 6). Presumably, determining the discourse focus means setting the discourse segment purpose as defined in Grosz and Sidner's (1986) theory of discourse. What it means to take the lead in developing the agents' plan depends on the plan representation but informally can refer to adding or taking away actions from the plan, rearranging actions, or setting parameters.

Whittaker and Stenton (1988) do not define initiative beyond calling it control of the dialogue by its participants. Their work is notable in that

¹The page numbers come from the digital version: <http://citeseer.nj.nec.com/244268.html>

they define a set of rules (see Figure 1) specifying who has initiative for each turn in a dialogue. These rules approximate the more complex definition given by Chu-Carroll and Brown and have been used in several projects because they facilitate reliable annotation (Strayer and Heeman, 2001; Jordan and Di Eugenio, 1997; Doran et al., 2001; Walker and Whittaker, 1990).

2.2 Initiative in human-human corpora

Previous work has shown a pattern to how initiative shifts among dialogue participants in problem-solving dialogues. Guinn (1996) used simulated conversational agents to argue that the most efficient problem-solving dialogues are those where the participant who knows the most about the current subtask takes initiative. The corpus analysis of Walker and Whittaker (1990) gives evidence that in natural dialogue, knowledgeable speakers do take initiative. Walker and Whittaker studied task-oriented dialogues (TODs) involving an expert guiding a novice through assembling a water pump, and advisory dialogues (ADs) involving an expert giving advice about financial and software problems. In the TODs, as we would expect, the expert had initiative most of the time (91% of the turns). However, ADs have closer to an equal sharing of initiative — the expert had initiative for 60% of the turns in finance ADs and 51% of the turns in software ADs. This is because in the ADs, the novice must communicate the details of his problem to the expert as well as the expert telling the novice what to do.

Shah (1997) investigated initiative in tutorial dialogue, typed human-human tutoring dialogues dealing with the circulatory system. Her corpus consisted of students' initial tutoring session and a subsequent session with each of the same students. She categorized student initiatives based on their communicative goal (*e.g.*, challenge, support, repair, request information). Shah found that the initial sessions had twice the number of student initiatives as the subsequent sessions. The nature of student initiatives also changed over time: the proportion of student initiatives associated with confusion (long pauses and self repairs) decreased in subsequent sessions and the proportion of challenges increased. Shah also looked at tutor reac-

tions to student initiatives; she found that tutors sometimes rejected student initiatives, but she did not investigate what triggered such actions.

Graesser and Person (1994) labeled student questions (a subset of the initiatives studied by Shah) in a corpus of tutoring sessions for a research methods course. Graesser and Person developed a taxonomy of different question types. Of specific interest are deep-reasoning and knowledge deficit questions. Deep-reasoning questions involve causal reasoning and hypothetical situations. Knowledge deficit questions are triggered when a student realizes an inconsistency or gap in his understanding or gets stuck on a problem. Graesser and Person found that in the first half of the course there was a negative correlation between overall number of student questions and exam scores. In the second half of the course, there were positive correlations between exam scores and the proportion of student questions that were deep-reasoning questions and the proportion of student questions that were knowledge deficit questions.

Our study focused solely on initiative and did not address the difficult problem of categorizing question semantics. Initiative is a noisy measure of student participation. Shallow questions such as “What do I do next?” were treated the same as insightful questions such as “Is a load basically the opposite of a source?”. Despite this interference, we hypothesized that high levels of initiative would characterize students who took control of their learning and as a result scored well in the post experiment test.

3 Our Initiative Study

This section is a summary of our methodology and results. For more details or to download the corpus or annotation manual, consult the web page http://www.cogsci.ed.ac.uk/~jmoore/tutoring/BEE_corpus.html.

3.1 Method

The setting for this study is a course on basic electricity and electronics (BEE) developed with the VIVIDS authoring tool (Munro, 1994). Students read four textbook-style lessons and performed six labs using a circuit simulator with a graphical in-

terface. (Rosé et al., 2000) describes an experiment where students went through these lessons and labs with the guidance of a human tutor (the same one for the entire study). Before the lessons, students were given pretests to gauge their initial knowledge. After being tutored, students took the same tests again. We refer to the difference in their scores as learning gain. There were three sets of tutoring sessions (a session means all the dialogue between the tutor and one particular student): (1) the trial sessions where the tutor was not given any instructions on how to tutor [3 students], (2) the Socratic sessions where the tutor was instructed not to give explanations and to ask questions instead [10 students], and (3) the didactic sessions where the tutor was encouraged to give explanations and then probe student understanding with questions [10 students]. During these sessions, the student and tutor communicated through a chat interface. We will refer to the logs of this chat interface as the BEE dialogues.

In previous work (Core et al., 2002), we addressed the question of whether these Socratic and didactic dialogues were really Socratic and didactic. We used interactivity to approximate “Socraticness”, and showed that the Socratic dialogues were more interactive than the didactic dialogues. On average in the Socratic dialogues: a greater proportion of tutor utterances were questions (42% vs. 29%); the students produced a higher percentage of words in the dialogues (33% vs. 26%); and tutor turns and utterances were shorter. It is debatable whether this means the dialogues are really Socratic and didactic but it proves they reflect different tutoring styles which is sufficient for the purposes of this study.

Rosé et al. (2000) addressed the issue of whether the Socratic dialogues in this corpus were more effective than the didactic ones. They found a trend for Socratically tutored students to learn more, but additional data is needed to verify this trend. Chi et al. (2001) performed a similar study; in this case, no difference was found between the two tutoring strategies. However, Chi et al. noted that the didactic tutors sometimes inadvertently revealed answers to questions on the post-test (the test given after tutoring to measure how much was learned). So we cannot say anything conclusive

```

if turn = command then
  speaker has initiative
if turn = question then
  if (last_turn = question or
      last_turn = command) then
    listener has initiative
  else speaker has initiative
if turn = statement then
  if last_turn = question then
    listener has initiative
  else speaker has initiative
if turn = prompt then
  listener has initiative

```

Figure 1: Rules for Assigning Initiative about the effectiveness of Socratic tutoring in the BEE domain or Socratic tutoring in general.

3.2 Initiative Annotation Method

The two definitions of initiative we considered were that of Chu-Carroll and Brown (1998) and Linell et al. (1988). We felt that the extra granularity provided by Linell et al.’s initiative ranks would not be necessary and adopted Chu-Carroll and Brown’s definition. However, this definition makes reference to discourse focus without giving guidelines as to how discourse focus is to be recognized during annotation. For this reason, we used Whittaker and Stenton’s initiative assignment rules (1988) as an approximation to Chu-Carroll and Brown’s definition of (dialogue) initiative. We did not attempt to annotate task initiative, but mention this issue again in the conclusions.

We first give details of the initiative assignment rules and then come back to the issue of whether this was a valid choice. Before the rules can be applied, each turn in the dialogue must be classified into one of the following types based on its main purpose: **assertions** - declarative turns used to state facts, **commands** - turns intended to instigate action, **questions** - turns intended to elicit information, and **prompts** - turns not expressing propositional content (*e.g.*, “yeah”, “okay”).

We used the rules in Figure 1 to assign initiative. These are the same as the rules given by Whittaker and Stenton except that we make the assumption that a statement following a question responds to that question.

A benefit of this annotation scheme is that in our corpus the majority of turns can be automatically labeled: **questions** often ended in question marks; **commands** often started with verbs; a list of com-

mon **prompts** (“okay”, “yeah”) allowed most of these to be labeled, and **statement** could be used to label everything else.

We needed human annotators to correct the automatic labeling. One of the authors of the paper and another human annotator (not a project member) corrected the automatic annotations. The annotators had a reference manual and trained on trial sessions of the dialogues. To test inter-annotator reliability, the author and external annotator labeled the same 757 examples taken from non-training data; the resulting inter-annotator reliability as measured with the kappa statistic was 0.92. Generally, kappa values above 0.8 are considered acceptable.²

Although these initiative assignment rules allow reliable annotation and are easy to implement, the question remains whether they actually capture initiative. It is clear that commands and questions not following questions (*i.e.*, not clarification questions) set the discourse segment purpose (*i.e.*, take initiative). The contentious aspects of these rules are assuming that answers never take initiative and that questions following questions never take initiative. It is simple to construct counter-examples to these assumptions; however, the rules work well in practice. Walker and Whittaker (1990) showed that third person and one anaphora rarely crossed segment boundaries marked by initiative changes annotated with these guidelines.³

It may be the case that these annotation assumptions fail on selected examples. However, in eliminating the assumptions it is likely that we will introduce more errors than we correct. For example, it is clear that some answers take initiative; if a speaker asks “what time is it?” and the listener gives more information than the current time, then the listener has taken initiative. However, if the speaker asks “what causes current to flow?”, it is much more difficult to say which answers take initiative. Similarly, it is difficult to say when a ques-

²These guidelines are based on comments by Krippendorff (1980) as summarized in Carletta (1996). Krippendorff considered the case of two annotated variables. He said that comparisons were reliable when the kappas for those variables were above 0.8.

³In this study, hierarchical discourse segments were annotated using changes in initiative as a starting point; these changes were taken as marking either a segment endpoint or the beginning of a nested segment.

tion following a question takes initiative. Some factors are the content of the second question, how many times the first speaker has been interrupted, and the reaction of the first speaker. But it seems very difficult to define these factors more precisely and to define how they interact.

3.3 Initiative Analysis

Our first analysis was to measure the average percentage of turns for which students had initiative in the Socratic and didactic dialogues. The Socratic dialogues had 1547 turns, 2853 utterances, and 23,451 words while the didactic dialogues had 1378 turns, 2993 utterances, and 26,195 words. Surprisingly, students had initiative for fewer turns on average (10%) in the Socratic dialogues than in the didactic dialogues (21%).⁴ These results show that students did not take advantage of the fact that the Socratic dialogues were more interactive, and did not ask more questions; in fact students asked fewer questions in the Socratic condition. We noticed that many student questions in the didactic dialogues followed explanations, perhaps because the long explanations confused students.

We next tested the relationship between initiative and learning gain. Since Socratic and didactic dialogues also differ in interactivity, we tested the relationship between learning gain and the interactivity measures of average percentage of words and utterances produced by the student and average percentage of tutor utterances that were **questions**. Figure 2 shows this data; the top graph shows that initiative varies erratically as learning gain increases; there is no relationship (Pearson’s $r = -.0689$, $n = 23$, NS) between these variables. The same graph also shows average percentage of words produced by the student; this does have a relationship with learning gain (Pearson’s $r = 0.6$, $n = 23$, $p < 0.005$). The bottom graph shows the relationship between percentage of utterances produced by the student and learning gain (Pearson’s $r = 0.56$, $n = 23$, $p < 0.005$), and the relationship between average percentage

⁴To analyze significance, we looked at average percentage of expert initiative per session rather than per corpus. For the didactic dialogues, this average is 82% and for the Socratic dialogues it is 90%, a significant difference ($t = 2.26$, $df = 18$, $p < 0.05$ two-tailed).

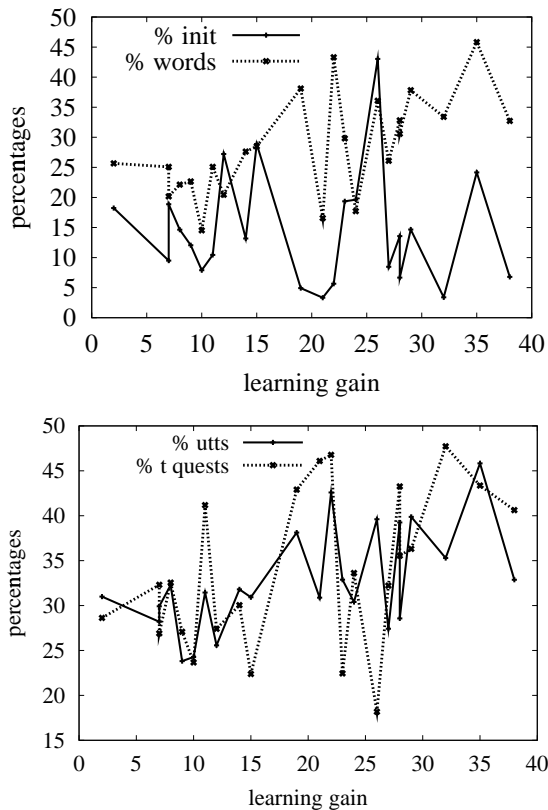


Figure 2: Learning Gain Comparisons

of tutor utterances that were **questions** and learning gain (Pearson's $r = 0.46$, $n = 23$, $p < 0.05$).

In section 1, we discussed the work of Walker and Whittaker (1990) on investigating initiative in the genres of advisory dialogues (ADs) and task oriented dialogues (TODs). Walker and Whittaker also investigated the difference between TODs in a spoken (telephone) modality and a typed (computer chat) modality. The results of their study are shown in columns 3-6 of Table 1 and the corresponding measures from our study are in columns 1 and 2. The Socratic dialogues have almost the same average expert initiative as TODs. In the TODs, the expert would issue a series of commands in order to get the novice to perform a procedure. In the Socratic dialogues, the tutor was issuing a series of questions in order to get the student to work through a line of reasoning to a correct answer.

The second row of the table shows average percentage of initiative changes that were abdications. Abdications are the use of prompts to give away initiative; these often occur after interrup-

tions⁵ to signal the original speaker to continue. Walker and Whittaker noted that spoken TODs had the most abdications but typed TODs had the least; modality has an impact on how initiative is managed.

In the didactic and Socratic dialogues (both of which are typed) shown in columns 1 and 2, we see that abdications are rarely used. A number of reasons are possible. In the typed TODs, communication consisted of two simultaneously updated channels. In the tutoring dialogues, participants would send each other short messages. This modality, typed text and restricted turn taking might have reduced the number of abdications. Another possible factor is that students in this study were relatively passive; the tutor could not rely on them to take initiative if she uttered a prompt. The tutor's initiative management also played a role. In our dialogues, after the student took initiative, the tutor would address the student's turn and then often take back initiative not giving the student a chance to utter a prompt.

4 Discussion

One interpretation of this data is that the definition of initiative was too crude and with a more precise definition, the results would show that students had more initiative in the Socratic dialogues than in the didactic dialogues. However, it would not involve simply changing three or four borderline examples. A large number of examples would have to change such that there was no longer significantly more student initiative in the didactic dialogues and instead significantly more student initiative in the Socratic dialogues.

A more likely interpretation is that when the tutor was employing the Socratic tutoring strategy, she did often take initiative (control of the dialogue) through constant questioning of the student. However, as shown by the interactivity statistics, students produced a higher percentage of words in the Socratic dialogues than in the didactic dialogues, and the percentage of words in the dialogue uttered by the student roughly correlated with learning. Given this correlation, we hypothe-

⁵Walker and Whittaker define interruptions as taking the initiative without invitation. It does not refer to interrupting the utterance of the other speaker.

	Didactic	Socratic	AD Finance	AD Software	TOD Phone	TOD Key
Expert-Initiative	79%	90%	60%	51%	91%	91%
Abdication	2.32%	0.43%	38%	38%	45%	28%

Expert-Initiative - % of total turns with expert initiative

Abdication - % of initiative shifts that are abdications

Table 1: **Initiative Measures for six Corpora**

size that student language production is an indication of student knowledge construction.

In future work, we see two ways of more closely measuring knowledge construction. The first is to use a question taxonomy such as (Graesser and Person, 1994) to identify deep tutor and student questions. (Jordan and Siler, 2002) suggests going further and classifying student answers. Although a tutor may ask a shallow question, the student may give more information than requested acting as if a deep question had been asked.

We plan to explore a second route based on discourse structure, in particular when a question has been dropped (*i.e.*, it has been answered correctly or abandoned). Our hypothesis is that in successful dialogues (ones where students learned the most), tutors do not drop questions until students correctly answer them meaning that the average discourse segment for a question is longer and may contain more nested segments.

5 Conclusions

In our corpus analysis, we found that initiative did not correlate with student learning and thus may not reflect activities such as problem solving and deep reasoning that lead to learning. Chu-Carroll and Brown (1998) identified the possibility that a speaker might have (dialogue) initiative but not be advancing the problem solving process. They created a measure called task initiative to track who is currently taking the lead in problem solving. For this measure to be useful in the tutoring domain, it will have to reflect student knowledge construction as well as problem solving participation. Our corpus analysis suggests that students may have such “learning” initiative without having dialogue initiative. We must further investigate this hypothesis in order to predict better the success of tutoring dialogues.

Our current results suggest that tutoring sys-

tems that encourage students’ language production will be most successful, and that a Socratic tutoring style is better at promoting student language production than didactic tutoring. These results may be good news for system builders; one possible Socratic teaching strategy would be to ask sequences of targeted questions where strong expectations about plausible answers make it easier to interpret student input.

However, we must be mindful of the fact that, even in Socratic interaction, students sometimes do take initiative rather than simply answering the sequence of questions posed by the tutor. It is not the case that human tutors simply brush off all student initiatives. And (Chi et al., 2001) shows that it is crucial that tutors do not plough ahead with their own plans, ignoring students’ signs of confusion. In future work, we will investigate the factors influencing the tutor’s decision about whether to entertain a student initiative, and investigate how these actions are signaled linguistically.

Acknowledgments

The research presented in this paper is supported by Grant # N00014-91-J-1694 from the Office of Naval Research, Cognitive and Neural Sciences Division. Thanks to Jean Carletta and our reviewers for their comments on this work.

References

- Jean Carletta. 1996. Assessing agreement on classification tasks: the Kappa statistic. *Computational Linguistics*, 22(2):249–254.
- Micheline T. H. Chi, Miriam Bassok, Matthew W. Lewis, Peter Reimann, and Robert Glaser. 1989. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13(2):145–182.
- Micheline T. H. Chi, Nicholas de Leeuw, Mei-Hung Chiu, and Christian Lavancher. 1994. Eliciting

- self-explanations improves understanding. *Cognitive Science*, 18(3):439–477.
- Micheline T. H. Chi, Stephanie A. Siler, Heisawn Jeong, Takashi Yamauchi, and Robert G. Hausmann. 2001. Learning from human tutoring. *Cognitive Science*, 25:471–533.
- Jennifer Chu-Carroll and Michael K. Brown. 1998. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User-Adapted Interaction*, 8:215–253.
- Jennifer Chu-Carroll and Jill S. Nickerson. 2000. Evaluating automatic dialogue strategy adaptation for a spoken dialogue system. In *Proc. of the 1st Annual Meeting of the North American Chapter of the ACL, Seattle*, pages 202–209.
- Mark G. Core, Johanna D. Moore, and Claus Zinn. 2002. Initiative in tutorial dialogue. In *Proc. of the ITS'02 Workshop on Empirical Methods for Tutorial Dialogue Systems*.
- Christine Doran, John Aberdeen, Laurie Damianos, and Lynette Hirschman. 2001. Comparing several aspects of human-computer and human-human dialogues. In *2nd SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark*.
- Barbara A. Fox. 1993. *The Human Tutorial Dialogue Project: Issues in the design of instructional systems*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Arthur C. Graesser and Natalie K. Person. 1994. Question asking during tutoring. *American Educational Research Journal*, 31(1):104–137.
- Arthur C. Graesser, Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9:495–522.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intensions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Curry I. Guinn. 1996. An analysis of initiative selection in collaborative task-oriented discourse. In *Proc. of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 278–285.
- Pamela W. Jordan and Barbara Di Eugenio. 1997. Control and initiative in collaborative problem solving dialogues. In *AAAI 1997 Spring Symposium on Computational Models for Mixed Initiative Interactions*, Stanford, CA.
- Pamela W. Jordan and Stephanie Siler. 2002. Control and initiative in computer-mediated human tutoring dialogues. In *Proc. of the ITS'02 Workshop on Empirical Methods for Tutorial Dialogue Systems*.
- Klaus Krippendorff. 1980. *Content Analysis: An Introduction to Its Methodology*. Sage Publications, Beverly Hills, CA.
- Per Linell, Lennart Gustavsson, and Päivi Juvonen. 1988. Interactional dominance in dyadic communication: a presentation of initiative-response analysis. *Linguistics*, 26:415–442.
- Douglas C. Merrill, Brian J. Reiser, and S. Landes. 1992a. Human tutoring: Pedagogical strategies and learning outcomes. Paper presented at the annual meeting of the American Educational Research Association.
- Douglas C. Merrill, Brian J. Reiser, Michael Ranney, and J. Gregory Trafton. 1992b. Effective tutoring techniques: Comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences*, 2(3):277–305.
- Allen Munro. 1994. Authoring interactive graphical models. In T. de Jong, D. M. Towne, and H. Spada, editors, *The Use of Computer Models for Explanation, Analysis and Experimental Learning*. Springer.
- Carolyn P. Rosé, Johanna D. Moore, Kurt VanLehn, and David Allbritton. 2000. A comparative evaluation of socratic versus didactic tutoring. Technical Report LRDC-BEE-1, University of Pittsburgh.
- Farhana Shah. 1997. *Recognizing and Responding to Student Plans in an Intelligent Tutoring System: CIRCSIM-Tutor*. Ph.D. thesis, Illinois Institute of Technology.
- John M. Sinclair and R. Malcolm Coulthard. 1975. *Towards an Analysis of Discourse: The English used by teachers and pupils*. Oxford University Press.
- Susan E. Strayer and Peter A. Heeman. 2001. Reconciling initiative and discourse structure. In *2nd SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark, September*.
- Kurt VanLehn, Stephanie Siler, Charles Murray, and William B. Baggett. 1998. What makes a tutorial event effective? In M. A. Gernsbacher and S. Derry, editors, *Proc. of the Twentieth Annual Conference of the Cognitive Science Society*. Erlbaum.
- Marilyn A. Walker and Steve Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proc. of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78.
- Steve Whittaker and Phil Stenton. 1988. Cues and control in expert-client dialogues. In *Proc. of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 123–130.