An Analysis of Human Tutors' Actions in Tutorial Dialogues

Vasile Rus, Nabin Maharjan, Lasang Jimba Tamang,
Michael Yudelson[1], Susan Berman[1], Stephen E. Fancsali[1], Steve Ritter[1]

Department of Computer Science/Institute for Intelligent Systems

The University of Memphis, Memphis, TN, USA

[1]Carnegie Learning, Inc., Pittsburgh, PA 15219
{vrus, nmharjan, ljtamang}@memphis.edu

Abstract

Understanding effective human tutors' strategies is one approach to discovering effective tutorial strategies. These strategies are described in terms of actions that tutors take while interacting with learners. To this end, we analyze in this paper dialogue-based interactions between professional tutors and tutees. There are two challenges when exploring patterns in such dialogue-based tutorial interactions. First, we need to map utterances, by the tutor and by the tutee, into actions. To address this challenge, we rely on the language-as-action theory according to which when we say something we do something. A second challenge is detecting effective tutorial sessions using objective measurements of learning. To tackle this challenge we align tutorial conversations with pre- and post- measures of student mastery obtained from an intelligent tutoring system with which the students interacted before and after interacting with the human tutor.

We present performance results of the automated tools that we developed to map tutor-tutee utterances onto dialogue acts and dialogue modes. We also report the most interesting emerging patterns in terms of tutor and tutees' actions. These patterns could inform our understanding of the tutoring process and the development of intelligent tutoring systems.

Introduction

A key component in tutoring is the use of effective instructional strategies, i.e. strategies that lead to students' learning gains. Discovering and validating such effective instructional strategies has been a key research goal in this area that was undertaken by many researchers (Aleven, Popescu, & Koedinger, 2001; Cade, Copeland, Person, & D'Mello, 2008; Jeong, Gupta, Roscoe, Wagster, Biswas, & Schwartz, 2008; Rowe, Mott, McQuiggan, Robison, Lee, & Lester, 2009). It should be noted that previous work distinguished between expert versus novice tutors while in our case we also distinguish between expert and effective tutors, as explained later.

The quest for effective strategies is even more critical and challenging given that average human tutors rarely enact sophisticated tutoring strategies (Graesser, D'Mello, & Person, 2009). Therefore, there is a need to discover and understand tutoring strategies that are either manifested by expert tutors, as opposed to novice or average tutors, or are motivated by pedagogical theory. In the latter case, the approach is to design theory-based strategies, implement them in an intelligent tutoring system (ITS; Rus, D'Mello, Hu, & Graesser, 2013), and then conduct controlled experiments to validate them. This "theory-based design and experimentation" approach has been adopted by a number of researchers (Graesser et al., 2001; Aleven, Popescu, & Koedinger, 2001; Rus, Banjade, Niraula, Gire, & Franceschetti, 2016).

The other approach, which we adopt here, is to discover strategies from expert tutors through manual or automated pattern analysis or data mining (DiEugenio, Kershaw, Lu, Corrigan-Halpern, & Ohlsson, 2006; Cade et al., 2008; Boyer, Phillips, Ingram, Ha, Wallis, Vouk, & Lester, 2011; Rus, Maharjan, & Banjade, 2015). An advantage of this data-driven approach to discovering tutorial strategies is that it allows researchers to take advantage of existing tutoring data collected from online human tutoring services.

Understanding what expert tutors do implies to, first, identify expert tutors and, second, develop a method to extract patterns of actions by the tutor and by the tutee that are associated with learning gains, which is our focus here, or other factors that impact learning such as learner's affect.

However, identifying expert tutors is non-trivial because tutoring expertise is yet to be understood (Rus, Maharjan, & Banjade, 2015). Human tutors may seem more expert than they actually are if, for example, they are very selective when it comes to their students, e.g. they may choose to work only with high-ability and highly-motivated students. On the other hand, a tutor who applies sound tutoring strategies may seem less of an expert and less effective if working with students who are low in ability and/or lacking motivation. Furthermore, it has been recently reported that experience or how much a tutor is paid, which have often been used as proxies for tutoring expertise, does not impact average learning gains nor does the tutor experience explain a significant portion of the variance in learning

gains (Ohlsson et al., 2007). It should be noted that Ohlsson and colleagues used a very small number of tutors in their study.

Consequently, we distinguish in our work between effective tutoring (the kind that induces learning gains) and expert tutoring (do the right thing, e.g. following sound pedagogical standards). It should be noted that this distinction is similar to research work in teacher expertise (Berliner, 2001) that distinguishes between good versus successful teachers: good teachers are those whose classroom performance meets professional teaching standards, whereas successful teachers are those whose students achieve set learning goals. It is beyond the scope of this paper to fully address the topic of tutoring expertise. Instead, we focus on effective tutoring by identifying effective tutors which in turn are identified by identifying effective tutorial sessions.

We have access to objective learning gains measures that allow us to identify effective tutorial sessions. In fact, we use a two-layer selection process to identify highly effective tutorial sessions. In a first layer, we select sessions from professional tutors, i.e. tutors who tutor to make a living. In a second selective layer, we use pre- and post-tutoring measures of mastery of target topics by aligning human tutorial sessions with sessions offered by Carnegie Learning's Cognitive Tutor (CT; Ritter et al., 2007), as explained later. Sessions that show high learning gains are deemed effective. We then analyze and compare effective versus less effective sessions in order to understand and characterize effective human tutoring. Furthermore, having access to pre- and post-tutoring mastery scores allows us to control for learners' prior knowledge, measured by pre-test mastery scores, when comparing effective versus less effective sessions. Controlling for students' prior knowledge is important for identifying truly effective tutoring, as highlighted earlier. This is, however, beyond the scope of this particular work reported here.

More specifically, we present in this paper our exploration of professional tutors' actions associated with effective tutoring sessions by analyzing human-to-human tutoring sessions provided by Tutor.com, a leading provider of human tutoring services. The main form of interaction in these tutorial sessions is chat-based conversation and therefore our focus is on analyzing dialogue-based tutoring sessions. Given that our data was collected from professional tutors, the results will be interpreted with this qualification in mind. There is no pre- and post-test when students interact with human tutors via chat which means we had to find a way to infer learning gains. In our case, the solution was to align the human tutoring data with another source of data, i.e. Cognitive Tutor data, from where learning gains could be derived. Students in our sessions are college-level, adult students who are required to interact with Cognitive Tutor and also have the option to ask for help from a human tutor. It is important to note that most students do not ask for help from a human tutor (Ritter et al., in press) which may imply a self-selection bias in our student population in the sense that it might consist of students that have higher meta-cognitive skills, e.g. they self-assess their knowledge and affective states and decide to ask for more help if needed, or prefer social interactions or appreciate affective support from a knowledgeable other human being. The results we present here should be interpreted with this important aspect of our data in mind.

Once effective sessions are identified, the second major step in the learning-from-expert-tutors approach, which in our case becomes learning-from-effective-tutors approach, is to characterize and explore tutors' actions and identify patterns of actions that are associated with learning gains. In our case, because we deal with dialogue-based tutorial interactions, first, we need to map the dialogue-based interactions, which are streams of utterances, into streams of actions. To this end, we rely on the language-as-action theory (Austin, 1962; Searle, 1969) to map speakers' utterances onto dialogue-acts. Dialogue acts are a linguistics construct that captures the general intent or action underlying a speaker's utterances. For instance, the intent or action behind the utterance "*Hello*" is to greet, similar to other utterances such as "*Good morning!*" or "*Welcome!*" In our case, all utterances are mapped into corresponding dialogue acts using, in our case, a predefined dialogue taxonomy (see details later). The taxonomy was defined by educational experts and resulted in a two-level hierarchy of 17 top-level dialogue acts and a number of dialogue subacts.

We adopted a supervised machine learning method to automate classify each utterance into one the dialogue act categories. It should be noted that other types of actions may be available to model student-tutor interactions, e.g. task actions as in Boyer and colleagues (2011), but in our case we only had dialogue interaction data.

Once tutorial dialogues were mapped onto sequences of dialogue acts, we were interested to identify chunks of actions that can be associated with general conversational segments and task-related or pedagogical goals. These chunks or segments are called dialogue modes. For instance, during a learner-tutor interaction it is fair to assume that there would be stretches of the interaction when the tutor would do more of the work by exemplifying and explaining the application of certain concepts, i.e. the tutor is modelling for the student the application of concepts, and therefore we call this part of the dialogue a *Modeling* mode. At other moments during the learner-tutor interaction, the roles would reverse with the student doing most of the work and the tutor only intervening when the student flounders, i.e. in this case the tutor scaffolds learner's application of concepts – we call such a segment of the dialogue a *Scaffolding* mode. We adopted a supervised method to automatically label dialogue modes in tutorial session. We basically learned from human-annotated data the signature of various dialogue modes using a sequence labeling framework based on Conditional Random Fields (CRFs; Lafferty, McCallum, Pereira, 2001).

Once tutorial sessions were mapped onto sequences of dialogue-acts and dialogue-modes, we analyzed the sessions in order to characterize what tutor and tutees do in effective sessions. As Rus, Graesser, and Conley (2014) noted, there could be only one tutoring strategy which is to make the learner apply effective learning strategies which, in turn, implies that we need to also analyze what tutees do in response to tutors' actions. We report our findings with respect to dialogue-act and dialogue-mode classification as well as the results of a number of session analyzed in terms of dialogue acts and modes.

Related Work

Discovering the structure of tutorial dialogues and tutors' strategies has been a main goal of the intelligent tutoring research community for quite some time. For instance, Graesser, Person, and Magliano (1995) explored collaborative dialogue patterns in tutorial interactions and proposed a five-step general structure of collaborative problem solving during tutoring.

Over the last decade, the problem has been better formalized and also investigated more systematically using more rigorous analysis methods (Cade, Copeland, Person, & D'Mello, 2008; Jeong, Gupta, Roscoe, Wagster, Biswas, & Schwartz, 2008; Chi, VanLehn, & Litman, 2010; Boyer, Phillips, Ingram, Ha, Wallis, Vouk, & Lester, 2011). For example, tutoring sessions are segmented into individual tutor and tutee actions and statistical analysis and artificial intelligence methods are used to infer patterns over the tutor-tutees action sequences. The patterns are interpreted as tutorial strategies or tactics which can offer both insights into what tutors and students do and guidance on how to develop more effective intelligent tutors that implement these strategies automatically. Our work contributes to this area of research by exploring tutors' actions by doing a tutorial data analysis at scale.

Language as Action

As pointed out earlier, in order to understand what tutors do we need to infer tutors' intentions and their general plan of action in the form of signature dialogue act mixtures and sequences, i.e. dialogue modes.

Speakers' intentions are modeled using elements from speech act theory (Austin, 1962; Searle, 1969). Speech act theory has been developed based on the language as action assumption which states that when people say something they do something. Speech act is a construct in linguistics and the philosophy of language that refers to the way natural language performs actions in human-to-human language interactions such as dialogues.

Its contemporary use goes back to John L. Austin's theory of locutionary, illocutionary, and perlocutionary acts (Austin, 1962). According to Searle (1969), there are three

levels of action carried by language in parallel. First, there is the locutionary act which consists of the actual utterance and its exterior meaning. Second, there is the illocutionary act, which is the real intended meaning of the utterance, its semantic force. Third, there is the perlocutionary act which is the practical effect of the utterance, such as scaring, persuading, and encouraging.

The notion of speech act is closely linked to the illocutionary level of language. Usual illocutionary acts are: greeting ("*Hello, John!*"), asking questions ("*Is it snowing*?"), making requests ("Could you pass the salt?"), or giving an order ("*Drop your weapon!*"). The illocutionary force is not always obvious and could consist of different components. As an example, the phrase "*It's cold in this room!*" might be interpreted as having the intention of simply describing the room, or criticizing someone for not keeping the room warm, or requesting someone to close the window, or a combination of the above.

A speech act could be described as the sum of the illocutionary forces carried by an utterance. It is worth mentioning that within one utterance, speech acts can be hierarchical, hence the existence of a division between direct and indirect speech acts, the latter being those by which one says more than what is literally said, in other words, the deeper level of intentional meaning. In the phrase, "*Would you mind passing me the salt?*", the direct speech act is the request best described by "Are you willing to do that for me?" while the indirect speech act is the request "*I need you to give me the salt.*" In a similar way, in the phrase "*Bill and Wendy lost a lot of weight with a diet and daily exercise.*" the direct speech act is the actual statement of what happened "*They did this by doing that.*", while the indirect speech act could be the encouraging "*If you do the same, you could lose a lot of weight too.*"

The present study assumes there is one direct speech act per utterance. This simple assumption is appropriate for automating the speech act discovery process. We do differentiate between top-level dialogue acts and second-level subacts but this is just a hierarchical organization of acts that allows us to analyze and process the dialogues at different levels of abstractness. A combination of an act and subact uniquely identifies, in this study, the direct speech act associated with an utterance.

The Dialogue Act and Mode Taxonomies

The current dialogue act taxonomy builds on an earlier version that was developed for a prior research project that sought to identify patterns of language use in a large corpus of online tutoring sessions conducted by human tutors in the domains of Algebra and Physics (Morrison et al., 2014). However, the taxonomy has been adapted to our new context; it is not identical to the one used by Morrison and colleagues (2014). The taxonomy is considerably more

granular than previous schemes such as the one used by Boyer and colleagues (2011).

The taxonomy employs two levels of description. At the top level, it identifies 17 standard dialogue categories including *Question*, *Answer*, *Assertion*, *Clarification*, *Confirmation*, *Correction*, *Directive*, *Explanation*, *Promise*, *Suggestion*, and so forth. It also includes two categories, *Prompt* and *Hint*, that have particular pedagogical purposes. Within each of these major dialogue act categories we identify between 4 and 22 subcategories. For example, we distinguish *Assertions* that reference aspects of the tutorial process itself (*Assertion:Process*); domain concepts (*Assertion:Concept*), specific approaches to the solution of a problem, such as the application of specific mathematical operations (*Assertion:Approach*); and the use of lower-level mathematical calculations (*Assertion:Calculation*). The taxonomy identifies 129 distinct dialogue act plus sub-act combinations.

The set of dialogue modes defined by the experts are: *Assessment, Closing, Fading, ITSupport, Metacognition, MethodID, Modeling, Off Topic, Opening, ProblemID, ProcessNegotiation, RapportBuilding, RoadMap, Scaffolding, Sensemaking, SessionSummary* and *Telling*. These modes are self-explanatory at some extent and, due to space reasons, we do not elaborate further.

Dialogue Act Classification

We assume that humans infer speakers' intention after hearing only few of the leading words of an utterance (Moldovan, Rus, & Graesser, 2011). One argument in favor of this assumption is the evidence that hearers start responding immediately (within milliseconds) or sometimes before speakers finish their utterances (Jurafsky and Martin 2009 - pp.814). This paper is yet another effort exploring the validity of such a hypothesis within the context of automated dialogue act classification of online chat posts.

Intuitively, the first few words of a dialog utterance are very informative of that utterance's dialogue act. For instance, *Question*s usually begin with a Wh-word while dialogue acts such as *Answer*s contain a semantic equivalent of yes or no among the first words, and *Greetings* use a relatively small bag of words and expressions.

In the case of other dialogue act categories, distinguishing the dialogue act after just the first few words is not trivial, but possible. It should be noted that in typed dialogue, which is less expressive than spoken dialogue, some information, such as intonation is lost. We should also recognize that the indicators allowing humans to classify dialogue acts also include the expectations created by previous dialogue acts. For instance, after a first greeting, another greeting, that replies to the first one, is more likely.

In the literature, researchers have considered rich feature sets that include the actual words (possibly lemmatized or stemmed) and n-grams. In almost every such case, researchers apply feature selection methods because considering all the words might lead to overfitting and, in the case of n-grams, to data sparseness problems because of the exponential increase in the number of feature values. Besides the computational challenges posed by such feature-rich methods, it is not clear whether there is need for so many features to solve the problem of dialogue act classification.

Therefore, we have selected first token, second token, third token, last token and utterance length as a set of features to represent a dialogue utterance. We did incorporate limited contextual clues in our experiments, e.g. the dialogue act of the previous utterance, as explained later.

Experiments and Results

Data: We used in our experiments a large corpus of 17,711 tutorial sessions between professional human tutors and college-level, adult students that was collected via an online professional human tutoring service. Students taking two college-level developmental mathematics courses (pre-Algebra and Algebra) were offered these online human tutoring services at no cost. The same students had access to computer-based tutoring sessions through Adaptive Math Practice, a variant of Carnegie Learning' Cognitive Tutor. A subset of 500 tutorial sessions containing 31,299 utterances was randomly selected from this large corpus for human annotation. The instances in the sample were randomly selected from the larger pool with the requirement that a quarter of these 500 sessions would be from students who enrolled in one of the Algebra courses (Math 208), another quarter from the other course (Math 209), and half of the sessions would involve students who attended both courses.

Expert Annotation Process: The session transcripts were manually annotated by a team of 6 subject matter experts (SMEs), e.g. teachers that teach the target topics, who were trained on the taxonomy of dialogue acts, subacts, and modes. Each session was first manually tagged by two independent SMEs without seeing each other's tags. Then, their tags were double-checked by a verifier, the designer of the taxonomy to resolve the discrepancies. The verifier had full access to the tags assigned by the independent SMEs. The average inter-annotator agreement was Cohen's kappa=0.72 for dialogue acts and kappa=0.60 for dialogue acts and subacts combined. The average independent annotator agreement for dialogue modes was kappa=0.38.

Dialogue Act and Dialogue Mode Classification: We built a classification model for predicting the dialogue act, dialogue act and subact, and dialogue mode labels, trained the model on the human annotated data, and then evaluated the trained model on a separate, unseen test data set using a 10-fold cross-validation approach. For space reasons, we summarize the results in terms of accuracy and Cohen's

kappa which indicates how well the output of our models agrees with the final tag adjudicated by the verifier while accounting for chance agreement.

We used Conditional Random Fields (CRFs) to tackle the dialogue act classification task. CRFs have several advantages over generative sequence labeling methods such as Hidden Markov Models (HMMs), e.g. CRFs models may account for the full context of a set of observations using features of various levels of granularity. Also, unlike other discriminative models such as Maximum Entropy Markov Models (MEMMs), CRFs do not suffer from the label bias problem.

Our CRF dialogue act model consists of the following features: the leading three tokens and last token from previous two, current, and next two utterances, current utterance length, previous dialog act, and bigram features composed of current first token - current second token, current second token - current third token and the trigram consisting of first token, second token and third token of current utterance. We also developed HMM model for classifying dialogue acts. The best results were obtained with CRFs as can be seen in Table 1.

Classifier	Accuracy (%)	Kappa
HMM	67.9	0.591
CRF	**74.3**	**0.671**

Table 1. Performance of dialogue act classifiers.

We also applied a number of other machine learning algorithms including Naïve Bayes, Decision Trees, and Bayes Nets. But we have not noticed any improvement with these different approaches. For space reasons, we do not show results for dialogue subact classification.

Dialogue mode labelling has been tackled both as classification task as well as a sequence labelling task using CRFs. The best CRFs-based model yielded an accuracy of 51.7% and kappa=0.48. The kappa for the dialogue mode labeling is better than the human inter-annotator agreement of 0.38.

Tutorial Session Analysis: Once all the sessions were mapped onto streams of dialogue acts and modes we proceeded with understanding the general structure of such sessions and identifying patterns of actions that are linked to learning gains. Learning gains were measured using several metrics generated by Cognitive Tutor (CT) such as number of assists per minute or per step. The number of assists measures the level of help a student needs while learning with the help of CT. We obtained the level of help a student needed in the CT session right before the human tutor session as well as the level of help needed in the CT sessions right after the human tutor session. A drop in the level of help needed is considered as evidence of progress or learning gains. An additional level of complexity to our analysis is added by the fact that sometimes the before and

after CT sessions may not be on the same topic. We differentiated in our analysis between human tutoring sessions that are in between Cognitive Tutor sessions that tackle the same topic or different topics.

As the next step, we conducted a series of comparison analyses of the human tutoring sessions' profiles in terms of dialogue act and dialogue mode distributions. We present only some of the analyses and findings due to space constraints. More specifically, in one type of analysis, we compared the profiles of the top 25% versus the bottom 25% sessions in terms of learning gains. Figure 1 shows an example comparison of two dialogue mode profiles corresponding to top 25% versus bottom 25% of sessions when decreasingly ranked based on learning gains measured as number of assists per minute. The figure shows the distributions of dialogue modes triggered by the tutors. A closer analysis of the two profiles revealed that in the top sessions there are relatively more *Fading* and *Telling* modes triggered by tutors, on average, and relatively more *Scaffolding* modes by tutors and less *Sensemaking,* on average. In the bottom sessions, there are relatively more *ITSupport* modes initiated by student and relatively more *ProcessNegotiation* modes initiated by both conversational partners.

We also did a quantitative comparison of the top and bottom sessions' profiles of dialogue acts and dialogue modes using Kullback-Leibler divergence and Information Radius. It should be noted that other comparison of this kind have been conducted which we do not present of space reasons.

The profile comparison offers a good way to compare the general mix of dialogue acts and dialogue modes of top sessions versus bottom sessions. However, they do not capture sequential information.
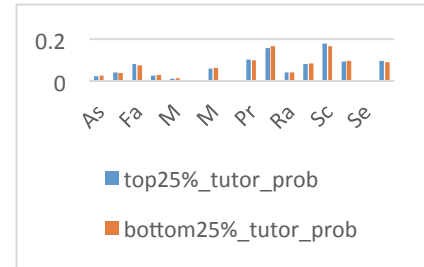

Figure 1. Dialogue mode profiles of top versus bottom 25% sessions, respectively.

To get a profile of a session that also captures the sequential information of dialogue acts and dialogue modes we used sequence logos, which can be used as an efficient visualization tool to represent distribution of various observations over discrete time. They are used in biomedical research to, for instance, visual genomic information such as sequences of genes. Figure 2 shows a sequence logo for tutorial sessions in terms of dialogue mode sequences. The logo regards each dialogue session as a discrete sequence of dialogue modes and then determines the dominant mode

at each discrete moment in the sequence. The dialogue mode at the top of a stack of modes at each discrete moment of the dialogue is the most frequent mode at that moment. Furthermore, the height of each letter in a stack represents the amount of information contained. The bigger the letter/mode at a particular discrete time the more certain the dominance of the corresponding mode is. For instance, at the discrete time 1 in the sequence logo shown in Figure 2 the dominant mode is *Opening*.
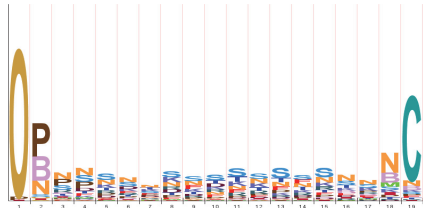

Figure 2. Dialogue mode sequence logo for sessions of average length 19.

From the sequence logo, we can infer the most certain sequence of dialogue modes in a typical human tutoring session as the sequence of the most certain dialogue modes at each discrete moment: O, P, N, N, S, N, N, S, S, S, S, S, S, S, S, S, N, N, N, C, where O – *Opening*, P – *ProblemIdentification*, N - *ProcessNegotiation*, K – *Sensemaking*, S – *Scaffolding*, T – *Telling*, F – *Fading*, C – *Closing*. This sequence of most certain dialogue modes can be regarded as a good overall summary of the tutorial sessions across all tutors, all students, and all topics. This summary was obtained by analyzing all sessions having 19 dialogue modes, which is the average length of a human tutoring session in terms of number of modes. As can be noted, the typical sequence of tutorial strategies is dominated by *Scaffolding*. This could be a consequence of the nature of the human tutoring sessions that we used which is mostly in the form of one-time interaction focusing on, mostly, homework help as opposed to a longer term tutor-tutee relationship spanning many sessions over a longer period of time.

Conclusion

We presented in this paper our approach to characterizing the human tutorial process which relies on analyzing tutorial sessions in terms of actions by the tutor and by the students. We used learning gains derived from students' interaction with a computer tutor and then conducted a profile and comparison analysis of the top and bottom, in terms of learning gains, human tutorial sessions. As part of our future work, we plan to conduct further analyzes while accounting for other factors such as students' prior knowledge.

References

Aleven V., Popescu, O., & Koedinger, K. R. (2001). Towards Tutorial Dialog to Support Self-Explanation: Adding Natural Language Understanding to a Cognitive Tutor. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), Proceedings of AI-ED 2001 (pp. 246-255). Amsterdam, IOS Press

Austin, J. L. (1962). How to do things with words: Oxford University Press, 1962.

Boyer, K. E., Phillips, R., Ingram, A., Ha, E. Y., Wallis, M., Vouk, M., & Lester, J. (2011). Investigating the relationship between dialogue structure and tutoring effectiveness: a hidden Markov modeling approach. *International Journal of Artificial Intelligence in Education*, *21*(1-2), 65-81.

Graesser, A. C., D'Mello, S. K., and Person, N., (2009). Meta-knowledge in tutoring.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in educational theory and practice*. Routledge.

Jurafsky, Dan.; and Martin, J.H. (2009). Speech and Language Processing. Prentice Hall.

Moldovan, C., Rus, V., & Graesser, A.C. (2011). Automated Speech Act Classification for Online Chat, The 22nd Midwest Artificial Intelligence and Cognitive Science Conference

Morrison, D. M., Nye, B., Samei, B., Datla, V. V., Kelly, C., & Rus, V. (2014). Building an Intelligent PAL from the Tutor.com Session Database-Phase 1: Data Mining. The 7th International Conference on Educational Data Mining, 335-336.

Ohlsson, S., DiEugenio, B., Chow, B., Fossati, D., Lu, X., and Kershaw, T.C (2007). Beyond the code-and-count analysis of tutoring dialogues. In AIED07, 13th International Conference on Artificial Intelligence in Edu-cation, Marina Del Rey, CA, July 2007.

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007) The Cognitive Tutor: Applied research in mathematics education. Psychonomics Bulletin & Review, 14(2), pp. 249-255.

Ritter, S., Fancsali, S., Yudelson, M., Rus, V., & Berman, S. (in press). Toward Intelligent Instructional Handoffs Between Humans and Machines, Workshop on Machine Learning for Education, The Thirtieth Conference on Neural Information Processing Systems (NIPS) 2016.

Rus, V., D'Mello, S., Hu, X., & Graesser, A.C. (2013). Recent Advances in Conversational Intelligent Tutoring Systems, AI Magazine.

Rus, V., Graesser, A.C., & Conley, M. (2014). The DENDRO-GRAM Model of Instruction, Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2), (Eds. Sottilare, R.), Army Research Lab.

Rus, V., Maharjan, N., & Banjade, R. (2015, June). Unsupervised discovery of tutorial dialogue modes in human-to-human tutorial data. In *Proceedings of the Third Annual GIFT Users Symposium* (pp. 63-80).

Rus, V., Banjade, R., Niraula, N., Gire, E. & Franceschetti, D. (2016). A Study On Two Hint-level Policies in Conversational Intelligent Tutoring Systems, The 3rd International Conference on Smart Learning Environment, 2016.

Searle, J.R. (1969). Speech Acts. Cambridge University Press, GB, 1969.