Automated Labelling of Dialogue Modes in Tutorial Dialogues

Vasile Rus, Nobal B. Niraula, Nabin Maharjan, and Rajendra Banjade

Department of Computer Science and Institute for Intelligent Systems

The University of Memphis, Memphis, TN, 38152, USA

{vrus, nbnraula, nmharjan, rbanjade}@memphis.edu

Abstract

We present in this paper a study whose goal was to automatically label higher level constructs, called dialogue modes, in tutorial dialogues. Each tutorial dialogue is regarded as a sequence of utterances articulated by either the learner or the tutor. The dialogue utterances can be grouped into dialogue modes which correspond to general conversational phases such as dialogue openings, e.g. when the conversational partners greet each other, or serve specific pedagogical purposes, e.g. a scaffolding students' problem solving process. Detecting dialogue modes is important because they can be used as an instrument to understand what good tutors do at a higher level of abstraction, thus, enabling more general conclusions about good tutoring. We propose an approach to the dialogue mode labeling problem based on Conditional Random Fields, a powerful machine learning technique for sequence labeling which has net advantages over alternatives such as Hidden Markov Models. The downside of the Condition Random Fields approach is that it requires annotated data while the Hidden Markov Models approach is unsupervised. The performance of the approach on a large data set of 1,438 tutoring sessions yielded very good results compared to human generated tags.

Introduction

A key research question in intelligent tutoring systems (Rus, D'Mello, Hu, & Graesser, 2013) and in the broader instructional research community is understanding what expert tutors do. This goal is motivated by research showing that expert tutors are more effective when it comes to student learning gains (2-sigma effect size which is equivalent to 2 letter grades improvement, e.g. from C to A) than unaccomplished tutors (effect size=0.4; Bloom, 1984).

Indeed, understanding what expert tutors do has been a research goal undertaken by theoreticians and empiricist alike (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Di Eugenio, Kershaw, Lu, Corrigan-Halpern, & Ohlsson, 2006; Cade, Copeland, Person, & D'Mello, 2008; Jeong, Gupta, Roscoe, Wagster, Biswas, & Schwartz, 2008; Boyer, Phillips, Ingram, Ha, Wallis, Vouk, & Lester, 2011; Lehman, D'Mello, Cade, & Person, 2012). A typical operationalization of this goal of understanding of what good tutors do was to define the behavior of tutors based on their actions. To this end, the learner-tutor interactions were broken down into primitive actions and then significant differences between expert tutors and less accomplished tutors are reported. For instance, Boyer and colleagues (2011) modelled the learner-tutor interaction as sequences of task actions (e.g., opening a file) and dialogue acts, i.e. actions behind utterances, while Cade and colleagues (2008) used just dialogue acts to model the learner-tutor interaction.

In our case, we model tutorial dialogues as dialogue-act sequences because there are no other types of actions, e.g. task actions as in Boyer and colleagues (2011), that we consider in our analysis. Our view of a tutorial dialogue as a sequence of actions is based on the language-as-action theory (Austin, 1962; Searle, 1969). According to the language-as-action theory, when we say something we do something. Therefore, all utterances in a tutorial dialogue are mapped into corresponding dialogue acts using, in our case, a predefined dialogue or speech act taxonomy. The taxonomy was defined by educational experts and resulted in a two-level hierarchy of 15 top-level dialogue acts and a number of dialogue subacts. The exact number of subacts differs from dialogue act to dialogue act. The overall, two-level taxonomy consists of 126 unique dialogue-act-subact combinations (Morrison, Nye, Samei, Datla, Kelly, & Rus, 2014). It should be noted that automatically discovered dialogue act taxonomies are currently being built (Rus, Graesser, Moldovan, & Niraula, 2012) but it is beyond the scope of this paper to automatically discover the dialogue acts in our tutoring sessions.

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 *modelling* 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. the tutor scaffolds learner's application of concepts process; in this case we would label such a segment of the dialogue a *scaffolding* mode. Discovering such dialogue modes automatically could be extremely useful for understanding what exactly good tutors do and transfer that understanding to the development of ITSs. Ideally, the dialogue modes should be discovered automatically based on actual learner-tutor sessions similar to the dialogue-act discovery process proposed by Rus and colleagues (2012). The alternative is to have a set of predefined modes proposed by experts, which is the framework adopted in this paper. That is, we focused on labelling tutorial dialogues with a set of expert-defined dialogue modes similar to the modes defined by Cade and colleagues (2008). We are exploring automated methods to discover modes, as in Boyer and colleagues (2011), but that work is in progress and is beyond the scope of this paper.

We adopted a supervised method in which we learned from human-annotated data the signature of various dialogue modes using a sequence labeling framework, i.e. Conditional Random Fields (CRFs; Lafferty, McCallum, Pereira, 2001). We report results with this approach on a *large data set* of 1,438 manually annotated sessions (95,526 utterances) of tutorial dialogues between human tutors and students. The data set is at least one order of magnitude larger compared to any other previous study, e.g. Cade and colleagues (2008) used 40 sessions. The results reported are very close to human experts.

The rest of the paper is organized as in the followings. The next section presents related work on dialogue act classification and dialogue mode discovery in tutorial dialogues. Then, we present our approach followed by the Experimental Setup and Results section in which we present the details of our experiments and results. We end the paper with a section on Discussion, Future Work, and Conclusions.

Related Work

An important task in dialogue-based educational systems is the detection of student intentions from their natural language input, i.e. utterances (Rus, Graesser, Moldovan, & Niraula, 2012). When the goal is to understand what tutors do, we also need to infer tutors' intentions and the general plan of actions in the form of signature dialogue act mixtures and sequences, which is the goal of this work. Speakers' intentions are modeled using elements from speech act theory (Austin, 1962; Searle, 1969). Speech act theory was developed based on the "language as action" assumption as explained later. Because we rely on dialogue acts to label the dialogue modes, we review next related work to the speech act theory, automated dialogue act classification, and dialogue mode identification.

Language as Action

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 consists 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 onesays 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. These simplifying assumptions are 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.

Automated Dialogue Act Classification

The task of speech act classification has been extensively addressed by the intelligent tutoring systems (ITS; Marineau, Wiemer-Hastings, Harter, Olde, Chipman, Karnavat, Pomeroy, Graesser, & the TRG, 2000; Serafin & Di Eugenio, 2004) and natural language processing (NLP; Reithinger & Maier, 1995; Ries, 1999; Stolcke, Ries, Coccaro, Shriberg, Bates, Jurafsky, Taylor, Martin, Van Ess-Dykema, & Meteer, 2000) communities. The related task of speech act prediction, which is about deciding what next speech act the automated dialogue system should generate, has also been investigated to some extent (Reithinger, 1995; Nagata & Morimoto, 1993; Bangalore & Stent, 2009).

In the automated speech act classification literature, researchers have considered rich feature sets that include the actual words (possibly lemmatized or stemmed) and n-grams (sequences of consecutive words). In our case, we adopted the approach in Moldovan, Rus, & Graesser (2011) and later extended by (Samei, Li, Rus, & Graesser, 2014). The approach is based on the observation that humans infer speakers' intention after hearing only few of the leading words of an utterance. 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).

Moldovan, Rus, & Graesser (2009) proved the validity of this hypothesis within the context of automated speech act classification of online chat posts. It should be noted that the focus of the paper is rather on dialogue mode labeling rather than classification. However, we needed dialogue acts for dialogue mode identification which is the reason we address this topic too.

Dialogue Mode Identification

Dialogue modes are sequences of dialogue acts that correspond to general conversational segments of a dialogue, e.g. an *Opening* mode corresponds to the first phase of the dialogue when the conversational partners greet each other, or to segments associated with pedagogical goals, e.g. a *Scaffolding* mode would correspond to the tutorial dialogue segment when the student works on something and the tutor scaffolds the learners' activity.

As already mentioned, previous research works on dialogue modes were based on both analytical and automated approaches. Cade and colleagues (2008) defined based on manual analysis a set of eight mutually exclusive tutorial modes: introduction, lecture, highlighting, modelling, scaffolding, fading, off-topic, and conclusion. An interesting aspect of their analysis is the granularity at which they defined the pedagogically important modes such as scaffolding, modelling, and fading. In their approach, these modes correspond to either the tutor or the students or both focusing on solving a full problem. In our approach, we used a different definition of modes proposed by Morrison and colleagues (2014). In this approach, a tutor or student could switch between proposed modes while working on a particular problem. That is, a particular mode is not associated with one problem solving task but rather with parts of such a problem solving task.

Boyer and colleagues (2011) used sequences of task actions and dialogue acts to automatically discover signature action sequences based on HMM model fitting. Furthermore, they related the automatically discovered modes to student learning. They discovered anywhere between 8 to 10 hidden states, or modes, depending on the tutor. Their set of modes includes: Correct Student Work, Tutor Explanations with Feedback, Tutor Explanations with Assessing Questions, Student Work with Tutor Positive Feedback, and Student Acting on Tutor Help. These modes are somehow different in their level of specificity from our dialogue modes. This is due mainly to the different process in which the modes were obtained (data-driven versus expert-defined), the fact that they also used task actions besides dialogue acts to discover the modes, and the fact that they discovered modes for an individual tutor whereas in our case we identify modes across many tutors.

Approach

Our approach to label dialogue modes follows a supervised machine learning approach in which we infer the

parameters of a model based on conditional random fields (CRFs; Lafferty, McCallum, Pereira, 2001). The trained model can then be used to label modes in new, unseen before tutorial dialogues.

Dialogue Act Taxonomy

The dialogue act taxonomy was developed with the assistance subject matter experts (SMEs), all experienced tutors and tutor mentors working for an online tutoring service, resulting in a fine-grained hierarchical taxonomy that includes 15 main categories. Each main dialog act category consists, in turn, of different subcategories resulting in an overall taxonomy of 126 distinct dialog act-subact combinations. Table 1 shows a list of main dialog acts with examples.

Act	Description	Example
Answer	A statement made in response to a question	Any non-zero integer.
Assertion	A free-standing statement (no prior question)	We have to keep the equation balanced.
Explanation	An utterance in the form of an explanation.	Because there are no horizontal forces acting on it.
Expressive	An utterance in the form of an expressive.	Oh!
Question	An utterance in the form of a question	What are you having trouble understanding?

Table 1. Examples of Dialogue Acts and corresponding utterances.

It should be noted that the dialog acts were defined and refined to minimize overlap between categories and maximize the coverage of distinct acts. The resulting taxonomy was described with examples and clear guidelines which were used by human annotators to tag the tutoring sessions; these manually labelled sessions were then used as training data in the supervised machine learning process based on CRFs. Details of the annotated data are given later.

Dialogue Mode Taxonomy

The set of dialogue modes defined by the experts are: Opening, Problem Identification, Assessment, Method Identification, Method Roadmap, Rapport Building, Process Negotiation, MetaCognition, Sensemaking, Fading, Scaffolding, Modelling, Telling, Session Summary, WrapUp/Close, and Off-topic. These modes are self-explanatory at some extent and, due to space reasons, we do not elaborate further.

Sequence Labeling with Conditional Random Fields

CRFs is a discriminative method for sequence labeling. It has several advantages over generative sequence labeling methods such as Hidden Markov Models (Rabiner, 1989), e.g. CRFs models may account for the full context of a set of observations using features of various levels of granularity, and over other discriminative models such as Maximum Entropy Markov Models (MEMMs), e.g. CRFs do not suffer from the label bias problem like MEMMs. CRFs have been successfully applied in a number of sequence labeling tasks such as POS-tagging (Lafferty, McCallum, Pereira, 2001), chunking (Sha & Pereira, 2003) and image segmentation (He, Zemel, Carreira-Perpinan, 2004). Therefore, CRFs are a best choice to label tutorial dialogues with dialog modes when expert-labeled data is available. As already mentioned, HMMs have the advantage of being unsupervised which means no expert-labeled data, which is expensive to obtain, is needed.

CRFs define conditional probability distributions $P(Y|X)$ of label sequences Y given input sequences X. CRFs can be viewed as a log-linear model for sequential labeling, i.e. a sequential version of logistic regression. That is, a log-linear model that relies on sum of weighted feature functions. Therefore, to use CRFs (for training or testing), we need to decide a set of feature functions. We defined our feature function as in the followings. For each utterance, we considered a context window of 5 utterances. We then computed feature functions by considering two immediate previous dialog acts and subacts (DA_{i-2}, DA_{i-1}, DSA_{i-2}, DSA_{i-1}), current dialog act and subact (DA_i, DSA_i) and two immediate next dialog acts and subacts (DA_{i+1}, DA_{i+2}, DSA_{i+1}, DSA_{i+2}). Next, we present our experimental setup and results when using this CRFs framework for dialogue mode labeling.

Experimental Setup and Results

Data

The data used in our experiments consisted of 1,438 sessions including 95,526 utterances generated by both tutors and students. We selected this data from a sample of 245,192 sessions obtained from an online tutoring service (Tutor.com). These sessions are about problem solving in the context of various Algebra and Physics topics. These are student-initiated sessions when they feel that they need help with solving a particular problem, e.g. as part of their homework.

The 1,438 sessions were manually annotated by a group of tutoring experts which were trained on both the dialogue act taxonomy and set of dialogue modes. When annotating independently, the inter-annotator agreement was 80.91% and kappa statistic was 0.77 for top-level dialogue acts and 64.90% and kappa of 0.63 for dialogue acts and subacts together. These values correspond to very good agreement

(kappa of 0.6-0.8 is considered very good agreement). For modes, agreement was lower: agreement of 55.03% and kappa of 0.47.

Dialogue Mode Identification

We present results with two different experiments. First, we report results when considering the golden, human-labeled dialogue acts and subacts annotated by the tutoring experts. Second, we automatically tagged each utterances with dialogue acts and subacts based on a dialogue act classifier we developed using a supervised machine learning approach following the basic idea in Moldovan, Rus, & Graesser (2011). The performance of this dialogue act classifier is very good at a kappa=0.71 for top level dialogue acts and kappa=0.49 for the full 126 dialogue act+subact taxonomy.

We used CRF++, a widely used and freely available tool for generating CRFs models for dialogue mode identification. We generated the input file for the tool from the annotated corpus in IOB2 format. Each instance consists of a sequence of utterances. Each utterance consists of four columns which are utterance id (combination of session id and utterance number), dialog-act, dialog-subact and the dialog mode.

We generated three different CRFs models by varying the set of features. In each experiment, we used the default values of the parameters and applied the 10-fold cross-validation to measure the performance.

The results corresponding to the experiments are presented in Table 2. In experiment 1, we only considered the dialog acts window (DA_{i-2}-DA_{i+2}) to generate the feature functions. The accuracy of the resulting model was 39.79%. Next, we incorporated dialog subacts features for experiment 2 (DSA_{i-2}-DSA_{i+2}). This increased the accuracy to 56.57%, surpassing the human agreement. The sharp gain in the accuracy suggests that the combination of dialog acts and subacts is very informative for predicting the dialog modes. In the third experiment, we enabled the bigram feature of the dialog modes (B) in addition to the features used in experiment 2. This improved the accuracy to 57.18%. There is just a small gain in accuracy when adding the bigram feature function B. Thus, we decided to use the model corresponding to experiment 2 for automatic classification of dialog modes in the unlabeled data.

Experiment	Feature Templates	Accuracy (%)
Experiment 1	DA_{i-2}-DA_{i+2}	39.79
Experiment 2	DA_{i-2}-DA_{i+2}, DSA_{i-2}-DSA_{i+2}	56.57
Experiment 3	DA_{i-2}-DA_{i+2}, DSA_{i-2}-DSA_{i+2}, B	57.18

Table 2. Summary of results for dialogue mode labeling. DA=dialog acts, DSA=Dialog Subact, B= Bigram.

In a second experiment, we repeated the above process except that we used automatically generated dialogue acts and subacts using the approach in Moldovan, Rus, & Graesser (2011) and Samei (2014). The performance was, as expected, lower at 28.77% accuracy. The major reason for the drop in performance of the dialogue mode labeling step based on CRFs is the presence of errors in the automatically generated dialogue acts and subacts, which are provided as input to the CRFs-based mode labeling module. Indeed, the accuracy of the dialogue act and subact classifier is good at 53% (kappa = 0.50) but there is room for improvement. On the other hand, it should be noted that the performance of the dialogue act and subact classifier approaches agreement of human experts (66%) when they annotated independently the utterances in the tutorial dialogues (Samei, 2014). The somehow good human expert agreement might indicate that the dialogue act and subact taxonomy is not operationally crisply defined as humans have some significant level of disagreements.

Discussion, Future Work, and Conclusions

We presented in this paper a CRFs-based approach to label dialogue modes in tutorial dialogues. The performance of the proposed approach surpasses human agreement when experts independently annotated modes. This is an outstanding result if taken at face value as many times human agreement is considered the ceiling of performance when it comes to certain tasks in particular in language and dialogue processing. It could be the case that the definition of dialogue modes is not crisp enough such that human annotators interpret differently the dialogue mode definitions. Automatically discovered modes using HMMs, which is our future work, will eliminate this problem.

The proposed method for dialogue mode labeling could be used as a building block to improve dialogue-based intelligent tutoring systems such as DeepTutor (Rus et al., 2013) or MetaTutor (Azevedo et al., 2009) and in the construction of a mixed system in which both automated and human tutors are accessible to a learner as envisioned by Rus, Conley, and Graesser (2014).

References

Austin, J.L. (1962). How to do Things with Words. Oxford University Press, 1962.

Azevedo, R., Witherspoon, A.M., Graesser, A.C., McNamara, D.S., Chauncey, A., Siler, E., Cai, Z., Rus, V., Lintean, M.C. (2009). MetaTutor: Analyzing Self-Regulated Learning in a Tutoring System for Biology, *Proceedings of the Internatioanl Conference on Artificial Intelligence in Education* (AIED 2009), July 6-10, 2009, Brighton, UK.

Bangalore, S., and Stent, A. (2009). Incremental parsing models for dialog task structure. In Proceedings of the Meeting of the European Chapter of the Association for Computational Linguistics.

Boyer, K.E., Phillips, R., Ingram, A., Ha, E.Y., Wallis, M.D., Vouk, M.A., & Lester, J.C. (2011). The International Journal of Artificial Intelligence in Education (IJAIED), Vol. 21 No. 1, 2011, 65-81.

Cade, W., Copeland, J., Person, N., & D'Mello, S. (2008). Dialogue modes in expert tutoring. In Woolf, B., Aimeur, E., Nkambou, R., & Lajoie, S. (Eds.) Proceedings of 9th International Conference on Intelligent Tutoring Systems (pp. 470--479). Springer, Berlin.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., and Hausmann, R. G. (2001). Learning from human tutoring. Cognitive Science, 25 (2001) 471–533.

Di Eugenio, B., Kershaw, T. C., Lu, X., Corrigan-Halpern, A., and Ohlsson, S. (2006). Toward a computational model of expert tutoring: a first report. 19th International conference of Florida Artificial Intelligence Research Society (2006) 503–508.

He, X., Zemel, R.S., Carreira-Perpinan, M. A. (2004). Multiscale conditional random fields for image labeling. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2004.

Lafferty, J., McCallum, A., Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. "Proc. 18th International Conf. on Machine Learning". Morgan Kaufmann. pp. 282–289.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments. In Intelligent Tutoring Systems: Vol. 5091. Lecture Notes in Computer Science (pp. 614-625). Montreal, Canada: Springer.

Lehman, B., D'Mello, S., Cade, W., & Person, K. (2012). How Do They Do It? Investigating Dialogue Moves within Dialogue Modes in Expert Human Tutoring. In. S. A. Cerri, & B. Clancey (Eds.), Proceedings of 11th International Conference on Intelligent Tutoring Systems (ITS 2012) (pp. 560-565). Berlin: Springer-Verlag.

Marineau, J., Wiemer-Hastings, P., Harter, D., Olde, B., Chipman, P., Karnavat, A., Pomeroy, V., Graesser, A., and the TRG. (2000). Classification of speech acts in tutorial dialog. In Proceedings of the workshop on modeling human teaching tactics and strategies at the Intelligent Tutoring Systems 2000 conference, pp. 65–71, 2000.

Moldovan, C., Rus, V., & Graesser, A.C. (2011). Automated Speech Act Classification for Online Chat, The 22nd Midwest Artificial Intelligence and Cognitive Science Conference, Cincinatti, OH, April 2011 (Best Student Paper Award - Honorary Mention).

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.

Nagata, M. and Morimoto, T. (1993). An experimental statistical dialogue model to predict the Speech Act Type of the next utterance. In Proceedings of the International Symposium on Spoken Dialogue (ISSD-93), pages 83--86, Waseda University, Tokyo, Japan.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

Reithinger, N. (1995). Some Experiments in Speech Act Prediction. In AAAI 95 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation, Stanford University.

Reithinger, N. and Maier. E. (1995). Utilizing statistical dialogue act processing in Verbmobil. In ACL95, Proceedings of the 33rd AnnualMeeting of the Association for Computational Linguistics.

Ries, K. (1999). HMM and Neural Network Based Speech Act Detection. In Proceedings of ICASSP 99.

Rus, V., Graesser, A., Moldovan, C., & Niraula, N. (2012). Automatic Discovery of Speech Act Categories in Educational Games, 5th International Conference on Educational Data Mining (EDM12), June 19-21, Chania, Greece.

Rus, V., D'Mello, S., Hu, X., & Graesser, A.C. (2013). Recent Advances in Conversational Intelligent Tutoring Systems, AI Magazine, 34(3):42-54.

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

Samei, B., Li, H., Rus, V., & Graesser, A.C. (2014). Context-based Speech Act Classification in Intelligent Tutoring Systems, The 12th International Conference on Intelligent Tutoring Systems, June 5-9, Honolulu, Hawaii.

Samei, B. (2014). Automated Speech Act Classification in Tutorial Dialogues, Masters Thesis, Department of Computer Science, The University of Memphis, Memphis, TN, November 2014.

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

Serafin, R. and Di Eugenio, B. (2004). FLSA: Extending Latent Semantic Analysis with features for dialogue act classification. ACL04, 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain.

Sha, F. and Pereira, F. (2003). Shallow parsing with Conditional Random Fields. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (NAACL '03). Morristown, NJ, USA: Association for Computational Linguistics, 2003.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics, 26(3):339–373.