RC24870 (W0910-053) October 9, 2009
Other

IBM Research Report

A Conditional Random Field Approach to Classroom Discourse Analysis Using Multilevel Features

Juan M. Huerta
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

A Conditional Random Field Approach to Classroom Discourse Analysis using Multilevel Features

Juan M. Huerta

IBM T J Watson Research Center
1101Kitchawan Road, Yorktown Heights, NY 10598
huerta@us.ibm.com

***Abstract*— In this paper we introduce a taxonomy of classroom discourse with particular focus on mathematical problem-solving discourse. We first discuss the hierarchical nature of classroom discourse and describe how our taxonomy addresses this hierarchical structure. We then describe an approach to classroom discourse classification based on our proposed taxonomy using Conditional Random Fields with features originating from multiple linguistic levels. The multilevel features reduce the classification error rate by over 40% compared with a purely unigram lexical features baseline. The framework and approach proposed in this paper can be useful in future work in education research, as well as discourse analysis research and intelligent tutoring applications.**

I. Introduction

In this paper we focus on the analysis of the verbal interaction that occurs in math classes between the teacher and the students when collaboratively addressing math problems. Classroom discourse is an important area in both dialog and education research as it provides important insights into collaborative problem solving, sheds light on the nature of dialog in general, as well as provides the basis for development education theories and support for education related applications.

In spite of its importance and its idiosyncratic characteristics, no specific taxonomy for the analysis of classroom discourse has been proposed so far in the computational linguistics and language processing community.

In this work we focus on two particular contributions: first, we propose a framework of analysis (i.e., a taxonomy of instructional discourse) motivated by the hierarchical nature of the classroom discourse, and second, we propose a computational approach to carry out discourse labelling using the proposed taxonomy based on Conditional Random Fields leveraging features originating from multiple linguistic levels. In our experiments we demonstrate the value of using these multilevel features.

As we will discuss in later sections, the proposed taxonomy is aligned with existing educational theories of instructional discourse, while at the same time is motivated by existing approaches to automatic analysis of dialog (specifically the attention/intention theory of discourse of Grosz & Sidner (G&S) 9]).

This paper is organized as follows: in section II we provide a background on recent and most relevant work that education researchers and dialog researchers have carried out in the area of dialog discourse analysis, and which serve as foundation to our taxonomy. In section III we focus on the nature of instructional discourse where we argue that the observed instructional discourse is driven by the goals and intentions of the teacher which provide its underlying hierarchical structure. In section IV we focus on the linguistic realization of classroom discourse, introduce our taxonomy of instructional discourse events, and elaborate how this taxonomy addresses the intention and attention focus described in section III. In section V we describe an approach to automatic discourse classification based on our proposed taxonomy using Conditional Random Fields (CRF) and multilevel linguistic features. In section VI we describe the classroom corpus we collected for the purpose of evaluation, the experiments we carried out and the results obtained. Section VII provides concluding analysis and discussion, including the potential impact of this work.

II. Previous Work on classrom discourse Analysis

In this section we provide a brief overview, as a way of background, on previous relevant research in classroom discourse. There are two main perspectives we consider: the educational perspective and the computational dialog research perspective [1]. The taxonomy that we introduce in later sections takes into consideration both perspectives.

From the education research perspective, because of the value of the pedagogical insight that collaborative classroom discourse provides, the analysis of classroom discourse has long played a vital role in education research. Many researchers in education have aimed at gaining pedagogical insight by observing the patterns of interaction and discourse in the classroom. For a perspective on the topic see Cazden, [5]. Educational studies are typically grounded in highly focused dialog analysis formalisms that are germane to the particular educational research framework used in each study. So broadly speaking there is no concurrence into a *single* discourse analysis framework in education. While going into details on diverse pedagogical theories and comparing their pros and cons is beyond the scope of the paper, we will choose to follow a specific pedagogical theory widely adopted especially in the area of mathematics education: the constructivism/scaffolding theory.

To briefly explain this theory, we focus on Meyer and Tuner [24]: their approach studies the phenomenon of

[1] By "computational" we mean approaches related to automatic natural language processing and understanding.

scaffolding and its relation with the self-regulation processes of students. Scaffolding is a process emanating from social constructivism learning theory (Hartman, [10]; Roehler et al, [27]) in which the learner is guided through the process of learning utilizing existing representations (knowledge) to create new representations. This theory is the cornerstone to broadly adopted current teaching practices and thus we will align our taxonomy to it. In Meyer and Turner's study, both teacher's scaffolding and student's self-regulation are assumed to be processes directly observable in the classroom discourse.

Scaffolding assumes a functional dualism of discourse (Werstch and Toma [30]) consisting of univocal and dialogical functions. In the univocal role discourse is meant to convey meanings directly and dialogic function refers on generating new meaning using existing meaning as basis. In our case *meaning* refers to mathematical *representations.* The Univocal-Dialogic dichotomy was later used in analysis approaches like the Transactive-Non-Transactive dichotomy used by Blanton [1], Huerta and Stylianou [12] and Huerta [11] which in turn generates from Kruger's work [16]. This functional dualism is reflected in the taxonomy introduced in section IV.

From the computational point of view previous relevant work has been divided in two. The first group makes analyses discourse without focusing particularly in classroom discourse (Ji [13], Zimmerman [31], Granell [8], Malioutov [20]). This work has led to the proposal of various approaches and taxonomies of dialog acts. Applications of these have included meetings and lectures (seminars). The second group has focused on educational discourse, but has looked at tutorial dialog especially with emphasis to computer-human interaction or one-on-one tutorial dialog (e.g., Litman [17,18], Purandare[25], Boyer [2,3,4], Liscombe[19], Sidner [28], Marineau [21]). The nature of one-on-one dialog and tutorial dialog is quite different from classroom discourse; however, some synergy between that body of research and our work can be attained.

Finally, text based approaches (to written discourse analysis) could be useful and relevant. For example Eisenstein [7] proposed an approach to hierarchical text segmentation. While speech and text are fundamentally different, the assumption of hierarchical structure that Eisner makes can be useful in the analysis of classroom discourse.

In the next section we further elaborate on why educational discourse is considered hierarchical and why the Grosz and Sidner framework presents an attractive foundation to our approach.

III. Hierarchical Discourse Analysis

In the previous section we identified the scaffolding (constructivist) perspective to classroom discourse analysis (from the educational viewpoint), and the Grosz and Sidner framework to dialog analysis from the computational perspective.

In this section we focus on the common assumption of these two perspectives: that the realization of discourse is the result of a hierarchical process. This assumption will allow us to present in later sections a taxonomy of classroom discourse that jointly considers the educational as well as the computational perspective.

From the educational viewpoint, high-level *curriculum* combined with lesson-level *goals* and plans jointly constitute the requirements that influence the actions taken by the teacher whom we consider to be the guide in the classroom dialog. Through scaffolding, the teacher guides the construction of representations utilizing existing representations based on strategies like direct lecturing, prompting etc., which are realized linguistically in clearly differentiated ways.

From the theory of G&S viewpoint, we associate teachers' discourse with a sequence of *intentions;* and the *attentional state* (focus of attention) to the artifacts, elements or objects related to a process of solving a mathematical problem. These artifacts are what we called *representations* in Section II.

In teacher-directed mathematical problem solving the dialog tends to center around properties of objects or processes that the teacher wants the students to learn, discover, or at least apply to a specific problem. In general these objects are the mathematical artifacts (formulae, tables, equations) and methods (proof, induction, solution of equation, equivalences, optimization etc) relevant to a mathematical discussion which map to particular instance solutions, their generalizations (formulas), their representation (algorithms), and the formalization of the solutions (proof). Problems and activities in a lesson plan are selected to highlight, teach or emphasize these artifacts and aspects and in a way to build these object into the student's own mental representations. These objects can also constitute methods and approaches to general problem solving that can cut across disciplines (e.g., induction, recursion, proof techniques, optimization, etc). For an interesting perspective in the topic see Reif [26]. These mathematical objects correspond to focus spaces. Together, the mathematical objects integrate focus structures.

For example, if the goal of a problem under discussion is to establish a generalized approach to a problem (in the form of a formula, for example) the object of attention are the aspects, variables, and components of this function or formula as it starts taking the shape of the solution. A teacher incrementally guides the classroom to provide more accurate and more generalized solutions to the problem by prompting, asking, emphasizing, explaining, etc.

In the Grosz & Sidner framework, attentional states refer to a group of focus spaces which together constitute a focus structure. Focusing, refers in this context to the process of elaborating and developing such mathematical representations. We can further elaborate in more detail on how specific concepts in the G&S theory map to classroom discourse (like ICP, OCP, DSP, DP etc), but we will leave this level of detail for future work.

IV. Linguistic Realization: A Taxonomy of Mathematical Classroom Discourse

Linguistic structure refers to the organization and relations of sequences of utterances that comprise discourse and discourse segments (Grosz & Sidner). In the Grosz and Sidner framework, as well as in some education research studies utterances in discourse are considered to "serve particular roles with respect to that segment". More specifically, Grosz and Sidner associate the roles of utterances in changing the intentional structure or the attentional state.

The questions we address now are: how do teachers in their role of ICP utilize these mechanisms? What is the nature of these utterance level mechanisms? Is it possible to devise a unified inventory of mechanisms?

If the repertoire of *roles* that a sentence can play in a discourse is relatively simple and small, these roles can be explicitly enumerated and computational approaches for their identification can be implemented. These approaches can leverage cue phrases if these constitute reliable features for classification.

Many education researchers have defined in multiple times taxonomies of discourse based on the roles the utterance play. The commonality in most of these studies is the characterization of the roles of classroom utterances into a simple set of well-defined roles. While not all these studies agree exactly on the specific set of roles existing in classroom discourse, a core set of commonly observed roles can be extracted. It also describes roles coming from both the observed (empirical) studies and the proposed (refined) patterns.

Education researchers have identified a family of patterns of interactions that constitute a common or pervasive technique - the IRE pattern, (Mehan, [23]): teacher initiation, student response and teacher evaluation.

Interestingly, this pattern (and small variations of it like the IRF pattern (Wells, [29]) in which the last step is teacher feedback) has been observed in classrooms from elementary levels (see section 5) to the college level (Blanton & Stylianou, [1]; Huerta [11]. Other researchers have looked into refinements or improvements in interaction strategies that could help improve learning (King, [14]). Some others have tried to understand interaction patterns with the goal to developing models for tutoring dialog and have not necessarily departed from the IRE pattern (for example, Marineau et al [20]; Boyer et al, [2-4]; Dzikovska et al, [6]).

Because of its pervasiveness we start by adopting the elements of the IRE/F as members of the instructional repertoire.

The first differentiation that we make in our taxonomy is the speaker: teacher and student. The second differentiation is the dichotomy of univocal and dialogic discourse (discussed in section II). For student discourse, the univocal discourse is ignored and one extra category is added (student question). Finally in addition to teacher univocal and dialogic discourse we create 4 additional teacher discourse tags that deal with low level mechanisms. We now describe the taxonomy.

The resulting set consists of Teacher's Discourse:

- Teacher Dialogic Transactive Prompt: The teacher's utterances are intended to promote transactive reasoning in the students and thus have mainly a dialogic role. Most of the times these are direct questions or prompts to the students. In the IRE pattern, this step corresponds to the Initiation Step.
- Teacher Dialogic Transactive Elaboration or Commentary: The teacher's utterances are a follow up or elaboration of a student's discourse sometimes producing a commentary.
- Teacher Univocal Instructive Didactive: Intended to convey knowledge and ideas where negotiation of the concepts are not intended or pursued directly. Mostly facts, problem statements, and lecturing.
- Teacher Coherent (Facilitative): The revoicing or implicit confirmation of the reception of a student idea. Most of the times repeating the words without emitting approval or disapproval. This is part of the E step in the IRE pattern.
- Teacher Explicit Affirmation or Negation (Facilitative): Because of its importance in the IRE/F pattern, we propose a category for the specific case where the teacher says "Yes" or "No" as a more specific case of the Teacher Directive case. This is also part of the E step in the IRE pattern.
- Teacher Logistic and Discipline Related: Gives exact instructions and commands to students, "Speak louder" "Stop playing" etc.
- Teacher Attention Redirection and Subject Identification
- Teacher Subjective Discourse or qualifier statement or emphasis: "That's an interesting idea". "I think you've been awfully quiet today". This is also part of the E step in the IRE pattern.

Student's discourse:

- Student Dialogic Response: The student's utterance, possible a response to a dialogic prompt whose goal is primarily dialogic. In the IRE pattern, this corresponds to the R step.
- Student Question: A specific and factual utterance of a student.

These utterance types have a dual role of affecting the attentional state and being used to drive the intentional state. For example, Teacher Dialogic Prompts utterances can be good candidates to denote an intentional shift in attentional focus. How effectively this is achieved depend among many other factors, in the skill and experience of the teacher, the state of the classroom, and the complexity of the material.

V. Discourse Classification based on Conditional Random Fields with Multilevel features

The taxonomy described above is meant to describe the most important discourse acts that comprise the dialog in a classroom. It generates from the scaffolding/constructivist educational perspective and is aligned with the intention level Sidner theory. We now propose a practical approach to classification using this taxonomy. We do not focus on discourse segmentation, but rather assume that the utterances that are to be classified correspond to a single class in our taxonomy. There is some discourse in which a single turns is comprised of multiple utterances, these sentence need to be broken down through segmentation, which is not the focus of this paper.

The taxonomy presented in the previous section can be the basis of educational research, but most importantly can constitute the foundation of computational applications, including intelligent tutors, classroom analysis etc. To be part of such systems our approach needs to work well with and leverage other components including ASR, parsers, etc.

Our approach for classroom discourse classification is based on conditional random fields [22]. To use Conditional Random Fields for classification we introduce features coming from various *linguistic* levels:

- Low level: diarization (i.e., speaker class id (student vs. teacher)), broad sentence class identification (e.g., question vs. nonquestion). This information is typically provided by hand by the transcriber or annotator or provided by an automatic component like the diarization component or speaker ID component. Giving us a queue of wether a student or an instructor is speaking, and wether the utterance is a question or not a question.
- Lexical features, syntactic structure, Name Entities: Words are transcribed by an ASR recognizer or transcribed by humans. These transcriptions can then be parsed syntactically. We focus strings of syntactic labels for each word. To build these, each label connecting a leaf (word) with the root of the tree is connected into a string. Additionally basic name entity is used (specifically students' names which is an important feature in support of redirection).
- Contextual: At this point we want to identify what that is currently being said was said verbatim in the previous turn by a different person. This essentially means to the identification of strings of Repeats of words with a threshold.

VI. EXPERIMENTS

We performed a set of experiments on data recoded in a middle school classroom. In this section we describe first some background into the nature of the data recorded, we then describe our corpus and then our experiments and results.

A. The Classroom Setting

Our corpus belong to a series of mathematic classes recorded over the course of three weeks in a New York City sixth grade classroom which was following the MiC (Math-in-Context) curriculum which is a comprehensive middle school mathematics curriculum for grades 5 through 8 developed by the Wisconsin Center for education research, (of the University of Wisconsin–Madison) and the Freudenthal Institute at the University of Utrecht, The Netherlands.The recorded classes address a series of problems focused on building and elaborating several algebra concepts. Our data was recoded using a portable video camera and the audio was extracted and transcribed by hand.

We find it useful now to provide a brief background in the education curriculum mandated at NY State Level in order to better understand the requirements under which a teacher operates and to better substantiate our claims made in sections II and III.

In our particular case, the lectures follow the NY State curriculum which is structured across 10 aspects (or "strands"). These strands can be divided in 5 Process Strands (Problem Solving, Reasoning and Proof, Communication, Connections, Representation), plus 5 Content Strands (Number Sense, Operation, Algebra, Geometry, Measurement and Probability). Lesson content is realized across these strands in the form of concrete curricula. Teachers are required to touch on these strands in their classes as they design and execute their lessons. At each school grade, there are objectives clearly defined in each of these strands. For example, a specific seven grade objective for the Communication Strand requires that the students learn to "provide an organized argument which explains rationale for strategy selection".

At this high level (i.e., curriculum and strands), a teacher might address several of these strands concurrently or individually in the course of a lesson or while discussing a problem in class. In guiding the classroom discussion, the teacher is expected to be mindful of the concrete goals in the curricular strands.

B. The Corpus

While we our recordings comprised classes spanning the course of 3 weeks, we specifically focused in the content of 4 lectures which together encompassed approximately 4 hours of recording. The data was manually transcribed. From these transcriptions, 65 segments were selected and extracted/

These 65 segments comprised the majority of the 4 hours of lectures, only leaving out non-interesting portions of the class. A non-interesting segment is defined as a segment without teacher transactive discourse or student participation. This type of non-interesting discourse amounted for less than 10% of the discourse of the 4 lectures.

Each of the 65 segments consists of a series of dialog turns which are topically coherent and typically gravitate around a focal point (which generally, the teacher is trying to make).

While the data was transcribed by hand, an Automatic Speech Recognition Component could be integrated and perform the transcription task. The quality of the recognition will greatly depend on the quality and location of the recording microphone(s) as well as the nature of the classroom environment (e.g., quiet vs. chaotic).

The resulting 65 segments comprised 1388 single-label turns out of which there were 1038 Teacher Turns and 350 Student Turns.

C. Corpus Annotation and Features

To annotate our data we focused on 7 Teacher discourse Labels and 2 Student Labels. Because of the selection process of interesting vs. non-interesting segments described above, none of our segments had the category of Teacher Univocal. A total of 11K Tokens (non-unique words) were in the corpus, out of which 7100 Teacher words (tokens) and 4650 Student tokens. Utterances were labelled by hand.

We then obtained the multi-level features described in section IV for each sentence: Words, Speaker Class, and Sentence Category were obtained from the transcription's data. We then performed Named Entity extraction and performed full syntactic parsing of the data using general English parser (Stanford Parser [15]). An informal review of the results of the parser indicates that while the out-of-the-box performance is quite reasonable, a very important direction of future work will be the identification of discourse structures emanating from mathematical representations (i.e., formulae and variables) can greatly enhance the quality of the parse and of the overall analysis. Finally we identified the coherent discourse by identifying repeated string of words in adjacent turns not belonging to the same person and ignoring stop words.

D. Classification Experiments

We performed a series of classification experiments. We are interested in labelling or classifying a sequence of observed discourse turns. These discourse turns consist of features which originate at various levels. We trained CRF models and tested classification accuracy in different feature conditions. To increase the level of statistical significance of the results, for each condition we ran a series of 14 trials. In each trial, we randomly partitioned the data set in approx 90% of the segments for training and 10% for testing then we trained CRF models and ran labelled and scored the test set. Our reported results are the average classification accuracy for the 14 trials in each feature condition. The CRF package we used was MALLET [22].

TABLE 1.
CLASSIFICATION ACCURACY RESULTS USING SEVERAL FEATURES

Features	**% Accuracy**
1-gram Lexical Features	48.85%
+Speaker Class	62.63%
+Sentence Class	67.49%
+Turn Coherency	71.17%
+Named Entity (Name)	72.16%
+Syntactic Label Strings	73.3%

Table 1 above shows the classification accuracy for each feature condition.

Our baseline consists on only utilizing 1-gram lexical features. The classification accuracy for this initial condition is slightly less than 50%. When we added the speaker class to the feature set the accuracy increased to 62%.

As we incrementally included the other features in the feature set (Sentence Class, Turn Coherency, Name Entity and Syntactic Label strings) the accuracy reaches up to 73.3%.

VII. DISCUSSION

In this paper we introduced a framework of discourse analysis dually motivated by the discourse theory of Grosz and Sidner as well as for the transactive framework emanating from education research.

The G&S framework assumes that the realization of the classroom discourse is the result of underlying intentions, which for the case of instructional discourse, are dominated by the instructor curricular and pedagogical goals and styles.

We identified a theory of mathematical classroom discourse –constructivism scaffolding- emanating from education research that aligns very well with the intention/attention assumptions of the G&S theory.

Based on these 2 theories we proposed a framework of discourse acts. While it is important to mention that Grosz and Sidner did not advocate for a fixed taxonomy of general dialog discourse, in our narrow domain we can bring forth a taxonomy without violating the assumptions of the theory of Grosz and Sidner given that the repertoire of mechanisms used in classroom discourse by teachers align well with G&S.

From the implementation point of view, we described an approach to classification of classroom discourse based on conditional random fields. We evaluated this approach under multiple features condition and demonstrated the value of incorporating features originating from various levels of linguistic analysis. The resulting classification error rate was reduced more than 40% from a words-only baseline as we introduced features of higher linguistic level. The final accuracy obtained is over 70% which indicates that this is a promising approach that could be useful in both computational applications as well as in support of education research (in support of computer assisted coding, for example).

We believe that our approach can be further refined as more features are incorporated into the feature set. We believe that promising features include mathematical name entity and mathematical entity coreference detection. These features should provide value in the classification task especially if the observed discourse deals with variable names and formulae. Providing a robust variable name coreference detection and detection prior to parsing is expected to help better parsing features. This, in turn, can result in higher classification accuracy and in turn better support in practical applications. Computational applications that can benefit from our classification approach include automatic tutors, automatic classroom discourse and lecture analysis and summarization etc.

We believe that our framework (or taxonomy) can be useful both education researchers and dialog researchers in analyzing classroom discourse. Education researchers can benefit from broader linguistic frameworks of analysis while dialog researchers can find in this area a rich source of research direction that provides insight to complex interaction mechanisms.

VIII. REFERENCES

[1] Blanton M., Stylianou D., David M., (2008) Understanding Instructional Scaffolding in Classroom Discourse on Proof, in The Learning and Teaching of Proof Across the Grades, Edited by D. Stylianou, M. Blanton, and E. Knuth, Routledge Publishers.

[2] Boyer K., Phillips R., Wallis M. Vouk M., and Lester J. (2009)The Impact of Instructor Initiative on Student Learning: A Tutoring Study. In *Proceedings of the 40th SIGCSE Technical Symposium on Computer Science Education (SIGSCE '09)*, pp. 14-18

[3] Boyer K., Phillips R., Wallis M., Vouk M., and Lester J.. (2008) Balancing Cognitive and Motivational Scaffolding in Tutorial Dialogue. In Proceedings of the Ninth International Conference on Intelligent Tutoring Systems, Montreal, Canada, pp. 239-249, 2008.

[4] Boyer, et al. Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models, Proceedings of NAACL HLT 2009: Short Papers, pages 49–52,

[5] Cazden, C. B. (2001). Classroom Discourse: The Language of Teaching and Learning. Portsmouth, NH: Heinemann. Second Edition.

[6] Dzikovska M., Callaway C.B, Stone M. and Moore J.D. (2006) Understanding student input for tutorial dialogue in procedural domains. BRANDIAL 06: Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue,

[7] Eisenstein, J. (2009). Hierarchical Text Segmentation from Multi-Scale Lexical Cohesion. NAACL 2009

[8] Ramon Granell, Stephen Pulman and Carlos D. Martinez-Hinarejos Simultaneous Dialogue Act segmentation and labelling using lexical and syntactical features SIGDIAL Workshop on Discourse and Dialogue (SIGDIAL 2009) Queen Mary University of London, September 11-12, 2009

[9] Grosz B., Sidner C., (1986) Attention, Intentions, and the Structure of Discourse. Computational Linguistics, 1986

[10] Hartman, H. (2002). Scaffolding & Cooperative Learning. Human Learning and Instruction New York: City College of City University of New York.

[11] Huerta J. M., (2009) "Discovering Discourse Motifs in Instructional Dialog", Tenth International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City.

[12] Huerta J. M., Stylianou D. (2008) "The Teaching Buddy: Speech and Language Technologies for Assisting and Assessing Instructional Practice", 7th European Conference on e-Learning, Cyprus

[13] Ji, Bilmes, (2005), Dialog Act Tagging Using Graphical Models, ICASSP 2005

[14] King Alison (1999) Discourse Patterns for Mediated Peer Learning, in Cognitive perspectives on peer learning, in Cognitive Perspectives on Peer Learning. Edited By Angela M. O'Donnell, Alison King Published by Lawrence Erlbaum Associates

[15] Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430.

[16] Kruger A. C. (1993). Peer Collaboration: Conflict, Cooperation, or both?. Social Development, 2, 165-182.

[17] Litman D., Moore J., Dzikovska M. O. and Farrow E.. (2009) Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains and Modalities. Proc. 14th International Conference on Artificial Intelligence in Education, Brighton, UK

[18] Litman and Kate Forbes-Riley Spoken Tutorial Dialogue and the Feeling of Another's Knowing SIGDIAL Workshop on Discourse and Dialogue (SIGDIAL 2009) Queen Mary University of London, September 11-12, 2009

[19] Liscombe J., Venditti J., and Hirschberg J. (2006), Detecting Question-Bearing Turns in Spoken Tutorial Dialogues, INTERSPEECH 2006, Pittsburgh

[20] Igor Malioutov; Regina Barzilay "Minimum Cut Model for Spoken Lecture Segmentation", in Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics

[21] Marineau, Wiemer-Hastings, Harter, Olde, Chipman, Karnavat, Pomeroy, Rajan & Graesser (2000) Classification of Speech Acts in Tutorial Dialogue. Workshop on modeling human teaching tactics and strategies (2000).

[22] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

[23] Mehan H., (1979) Learning Lessons, Cambridge MA. Harvard University Press.

[24] Meyer, D. ; Turner, J. (2002) Using Instructional Discourse Analysis to Study the Scaffolding of Student Self-Regulation Educational Psychologist 2002

[25] Purandare A. and Litman D.. (2008) Content-Learning Correlations in Spoken Tutoring Dialogs at Word, Turn and Discourse Levels. *Proceedings 21st International FLAIRS Conference*Reif F., (1995) "Understanding and teaching important scientific thought processes", *American Journal of Physics* 63 (1995) 17-35

[26] Reif F., (1995) "Understanding and teaching important scientific thought processes", American Journal of Physics 63 (1995) 17-35

[27] Roehler L.R., Cantlon D.J. (1997) "Scaffolding: A powerful tool in social constructivist classrooms "In K. Hogan & M. Pressley

[28] Sidner C., (2005) COLLAGEN and Tutoring Applications: Collaborative Discourse Theory in Tutorial Dialogues, University of Edinburgh, April 2005.

[29] Wells G., (1993) Reevaluating the IRF Sequence: A proposal for the articulation of theories activity and discourse for the analysis of teaching and learning in the classroom. Linguistics and Education, 5, 1-37.

[30] Wertsch, J Toma C.. (1995) Discourse and Learning in The Classroom: A sociocultural Approach. In Constructivism in Education, Edited by L. P. Steffe, J. E. Gale, Lawrence Erlbaum.

[31] Zimmerman, Stolke, Shriberg, Joint Segmentation and Classification of Dialog Acts in Multiparty Meetings, ICASSP 2006