Finding Transactive Contributions in Whole Group Classroom Discussions

Hua Ai, Marietta Sionti, Yi-Chia Wang, Carolyn Penstein Rosé, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, {iamhuaai,marietsionti}@gmail.com, {yichiaw,cprose}@cs.cmu.edu

Abstract: Automatic analysis of discussion logs from studies of computer supported collaborative learning has become more prevalent as technology for machine learning and text mining have become more powerful and usable. We extend this effort to analysis of transcripts of whole group classroom discussions. We argue in favor of adapting a transactivity based framework developed for analysis of collaborative learning discussions for use in the context of whole class discussions. Results from an evaluation of this technology applied to transcripts from 3 class periods of middle school math demonstrate promising results, specifically Cohen's Kappa of .69 in comparison with human annotators for identifying transactive contributions and .68 for identifying which prior utterance a transactive contribution relates to. Potential applications for teacher professional development, automatic assessment of learning processes, and real time support for classroom discussion facilitators is discussed. Implications for theory building related to learning in classroom discussions is also addressed.

Keywords: Corpus analysis, classroom discourse, automatic text processing techniques, argumentation

Introduction

In recent years the learning sciences community has seen a growing body of work towards automatic analysis of conversational interactions in learning contexts (Soller & Lesgold, 2000; Cakir et al., 2005; Erkins & Janssen, 2006; Rosé et al., 2008). Much of this work has taken place within the computer supported collaborative learning community (Wang et al., 2007; Kumar et al., 2007; Kumar et al., 2009), as part of an effort to develop dynamic support for collaborative learning that is capable of triggering interventions in a way that is responsive to what is happening within the conversation. In this paper we present an extension of automatic conversation analysis technology into a new area, specifically analysis of whole class group discussions. Our automatic analysis involves a construct referred to as Transactivity (Teasley, 1997; Weinberger & Fischer, 2006; Berkowitz & Gibbs, 1983), which is well studied in the domain of educational psychology and computer supported collaborative learning. Transactive contributions are arguments constructed in such as way as to reference, sometimes described as "operating on", the previously expressed reasoning of self or others.

Transactivity is an important construct to target for automatic analysis technology because of the extent of its prevalence within the literature on collaborative learning process analysis and the body of work that supports its value as a property of discussions for learning (Azimita &Montgomery, 1993; Joshi & Rosé, 2007). The body of work already demonstrating the feasibility of using machine learning technologies to identify transactive conversational moves in small group discussions (Rosé et al., 2008; Joshi & Rosé, 2007) offers hope that it can be extended into this new context. Ideas related to effective patterns of discussion in classroom contexts have evolved within their own separate history from that of the community of researchers studying analysis of collaborative learning interactions. Nevertheless, a growing subcommunity of the classroom discourse community has focused on facilitation strategies for group discussions that have very similar motivations relating to encouraging children to articulate their reasoning and to listen to and respond to the reasoning of others (Chapin & O'Connor, 2004; Resnick et al., 1992; Resnick, Michaels, & O'Connor, in press; Michaels, O'Connor, & Resnick, 2007). Similarly, within the problem based learning community, where discussion groups are smaller, but similarly lead by skilled facilitators, again similar ideas have emerged (Hmelo-Silver & Barrows, 2006). Thus, it is reasonable to expect that similar technology would be capable of identifying transactive moves in whole class discussions. Nevertheless, large group discussions are more complex than those that take place within small groups of two or three participants. Thus, extension of earlier automatic analysis technology to whole group discussion data is a research contribution in its own right.

In the remainder of the paper we describe our adaptation of the Berkowitz and Gibbs transactivity coding scheme for classroom discussion. We then describe experiments towards automatic analysis of transactivity from transcripts of classroom discussions. We then discuss potential applications of this technology for teacher professional development, automatic assessment of learning processes, and real time support for classroom discussion facilitators. Implications for theory building related to learning in classroom discussions is also addressed. We conclude with a discussion of remaining challenges for producing automatic transactivity analyses from classroom discussions.

Theoretical Framework

This paper presents our preliminary study on automatically analyzing transactivity in a classroom setting. This study is part of a larger effort related to building conversational agents that can assess the level of transactivity in ongoing discussions on the fly and respond accordingly. This work is based on the assumption that an assessment of level of transactivity in an in-progress discussion would be valuable information for facilitators to use in deciding what is needed from them to keep the conversation on a productive path in their role as orchestrators of group talk. This assumption is based on investigations of valuable conversational contributions in learning contexts that have been conducted both within communities exploring the cognitive foundations of group learning as well as the sociocultural community. Regardless of the theoretical framework, the same ideas have surfaced pointing to the value of transactivity or other very similar constructs as properties of conversation for learning. It should be noted that we are not arguing that this transactivity assessment is sufficient for facilitators to decide how to proceed, only that it provides useful insight into the level of productive engagement each student in the group is or has been exhibiting over the course of the conversation. Our work focuses on discussions involving young children, specifically middle school children learning math, although transactivity as a construct is not specific to math, and thus its definition does not refer to any specific math knowledge. Nevertheless, in order for the construct to be applied in its most valid sense, as we will see below, knowledge of math would be required in order to fully identify all of the connections between reasoning displays that students are making. Because of this, the approach we describe in the next section of this paper will be seen as necessarily a simplification of this construct. Nevertheless, the results show promise that although the annotations are being applied through automatic means, the reliability in comparison with human coding is high. Methodological implications of such simplifications are discussed at length by Rosé and colleagues (2008) in connection with automatic analysis of small group collaborative learning discussions.

The purpose of this section is to argue for the value of this transactivity analysis in connection with prior work in the learning sciences. For example, one cognitive justification for the value of transactive conversational behavior is its connection with cognitive conflict (Piaget, 1985). While we do not argue that transactivity is synonymous with cognitive conflict, we do argue that where transactive conversational moves highlight differences between the mental models of collaborating students, these moves provide opportunities for students to reflect on differences in their own mental models, which may then lead to cognitive conflict. One can argue that a major cognitive benefit of collaborative learning is that when students bring differing perspectives to a problem-solving situation, the interaction causes the participants to consider questions that might not have occurred to them otherwise. This stimulus could cause them to identify gaps in their understanding, which they would then be in a position to address. This type of cognitive conflict has the potential to lead to productive shifts in student understanding. It has the potential to elicit elaborate explanations from students that are associated with learning (Webb, Nemer, & Zuniga 2002). While transactivity has been valued previously for its role in consensus building discussions (Weinberger & Fischer, 2006), it should be noted that conversations exhibiting transactivity need not end in a common understanding or belief about an issue, and its value for learning can be seen as independent of whether this occurs or not. Furthermore, the value of transactivity is not limited to its potential cognitive benefits. From the sociocultural perspective, based on Vygotsky's seminal work (Vygotsky 1978), we can similarly argue that when students who have different strengths and weaknesses work together, they can provide support for each other that allows them to solve problems that would be just beyond their reach if they were working alone.

The most popular formalization of the construct of transactivity (Berkowitz and Gibbs, 1983) has 18 types of transactive categories, which characterize each child's conversational turn, as long as it is considered an explicit reasoning display. Before considering which of these codes, if any, is appropriate for a contribution, one must first determine whether that contribution constitutes an explicit articulation of reasoning, or at least a reasoning attempt. Beyond this, transacts have been divided in three types: elicitational, representational and operational, which ultimately were reduced to two, incorporating the elicitational in representational (R), which is considered a lower level transact, since it elicits or re-presents another's reasoning. On the contrary operational transacts (O) present a person's new argumentation, which is formed by building on another's contribution. A transact may also combine both types (R/O), because the boarders might be vague in some cases. The other two dimensions of transactive moves are focus and mode. Depending on the primary focus, a transact might be self-oriented (ego, operates on the speaker's own reasoning) or other-oriented (alter, operates on the reasoning of a partner, dyad shared opinion). Mode indicates if the transact was expressed competitively (i.e., the two expressions of reasoning are not consistent with one another) or non-competitively (i.e., the two displays of reasoning are consistent with one another). We have simplified the formulation considerably in this work and simply identify utterances according to whether they do not count as an articulation of reasoning, count as reasoning but not transactive, and transactive reasoning. In the case of transactive utterances, we identify which previous utterance is being related to.

In our formulation, articulation of reasoning by students is the goal, and thus we define what "counts" as a reasoning move. Our formulation of what counts as a reasoning display comes from the Weinberger &

Fischer (2006) notion of what counts as an "epistemic unit", where what they look for is a connection between some detail from a scenario (which in their case is the object of the case study analyses their students are producing in their studies) with a theoretical concept (which in their case comes from the attribution theory framework, which the students are applying to the case studies). When they have seen enough text that they can see in it mention of a case study detail, a theoretical concept, and a connection between the two, they place a segment boundary. Occasionally, a detail from a case study is described, but not in connection with a theoretical concept. Or, a theoretical concept may be mentioned, but not tied to a case study detail. In these cases, the units of text are considered degenerate, not quite counting as an epistemic unit.

We have adapted the notion of an epistemic unit from Weinberger & Fischer, rather than using it the same way Weinberger et al. did in their work, both because the topic of our conversations is very different in nature and because we're working with a much younger group of students. We consider that the basic requirements for a unit of talk to count as a reasoning display is that it has to contain evidence of a connection between some detail from the problem the students are trying to solve and some mathematical idea, which could be a theorem or an idea from an earlier problem they solved that they explicitly mentioned (because it shows evidence of making an abstraction), or some idea from a book that they explicitly mentioned. We would like to distinguish this from just parroting what they have heard. While teachers may use repetition to keep an idea salient within a discussion, which may be a very valuable facilitation move, we would rather not count places where students simply repeat what they have heard as evidence of reasoning. In our current formulation, we are considering the problem that the teacher has set up as what is given. We would like to make a distinction between what is given and what the students contribute beyond that.

So far we have located the instructor somewhat outside of the discussion the students are having, seeing the instructor as stimulus and support for the discussion and not actually part of the discussion. Note that we consider this scaffolding by the instructor as extremely valuable, and even necessary. However, we see it as distinct from the effect that teachers are attempting to achieve in the conversation between the students. Thus, in our current formulation, we are making a distinction between moves that facilitate the drawing out and refining of the kids' reasoning from the actual reasoning attempts. Thus, this paper only focuses on the analysis of the student utterances. As an additional caveat, while it was true also for analyses of adult discussions, it is even more true of these child discussions that we need to allow for displays of incorrect, incomplete, and incoherent reasoning to count as reasoning, as long as in our judgment we can believe an attempt at reasoning was going on. That will necessarily be quite subjective – especially in the case of incoherent explanations. However, we have achieved a human agreement of .68 Kappa using the definitions of articulation of reasoning and what it means to be a transactive utterance.

Machine Learning Technology

Machine-learning algorithms can learn mappings between a set of input features and a set of output categories. They do this by using statistical techniques to find characteristics of hand-coded "training examples" that exemplify each of the output categories. The goal of the algorithm is to learn rules by generalizing from these examples in such a way that the rules can be applied effectively to new examples. The goal is to replace this low level coding effort by humans rather than to perform an analysis over human coded data as has been done frequently in other work (e.g., Nathan, Kim, and Grant, 2009). In order for the machine learning process to work well, the set of input features provided must be sufficiently expressive, and the training examples must be representative. Typically, machine-learning researchers design a set of input features that they suspect will be expressive enough. At the most superficial level, these input features are simply the words in a document. But many other features are routinely used in a wide range of text-processing applications, such as word collocations and simple patterns involving part of speech tags and lexical features; we will draw from this prior work.

Once candidate input features have been identified, analysts typically hand code a large number of training examples. For example, tools such as the publically available TagHelper tool set (Rosé et al., 2008) have the capability of allowing users to define how texts will be represented and processed by making selections on the GUI interface. In addition to basic text-processing tools such as part-of-speech taggers and stemmers that are used to construct a representation of the text that machine-learning algorithms can work with, a variety of algorithms from toolkits such as Weka (Witten & Frank, 2005) are included in order to provide the means to learn mappings between the input features and the output categories.

Evaluation

We evaluated the use of machine learning technology for the analysis of transactive conversational moves in three classroom sessions where a teacher worked with a group of kids to elicit their reasoning about integers. Altogether the corpus covers 185 minutes of discussion involving the teacher and the students. The sessions have been transcribed. Altogether, the transcripts contain 707 student turns and 646 teacher turns. In this corpus, the students and the teacher contribute about equally.

Predicting Student Moves

Student moves were coded as Not Reasoning (NR), Externalization (E), or Transactive (T). Not reasoning is a large category that includes all student utterances that are off task, are related to management (e.g., responds to announcements), are plain answers to a teacher's question, or are simply repeating what had already been articulated. A detailed discussion can be found in (Sionti et al., 2010). Reasoning utterances are grouped into two categories. Externalization refers to students' reasoning that does not connect with any previously displayed reasoning. Transactive utterances are student utterances that display reasoning that connects with previous reasoning displays in some way. For example, in this dialog exerpt below, the first student utterance is tagged as Externalization while the second student utterance is tagged as Transactive.

> Student: I sort of agree but I also disagree. I wanted to say maybe the weight line shows you more because it shows like the materials and it also showed how much they weighed by how far apart they are...
> Teacher: Can you say a bit more about that?
> Student: Like the mineral oil is way heavier than the whatsit [looks at chart], the organic materials, but the mineral oil and the water, the fresh water are way closer. Like you could fit the length of 2 mineral oils in the length of 1 gravel.

Table 1: Student Moves in three classroom sessions

	T	E	NR
session1	75	41	112
session2	77	25	126
session3	90	25	136

In the case of transactive contributions, the prior utterance that the transactive utterance refers to, or operates on, is identified. Table 1 displays counts of student moves (**stuMov**) in each of the classroom sessions. We observe that students produced turns identified as articulations of reasoning (T or E) 47.1% of the time. Thus, a trivial classifier that simply assigns the category of NR, which was the most frequent category, would achieve a percent accuracy of 52.9%, although this corresponds to a Cohen's kappa of 0, which we will treat as our baseline for comparison. In the case of identifying which utterance a transactive utterance is responding to, again a simple heuristic achieves a moderate accuracy. Specifically, when predicting responding student utterance ID (**respondUtt**), we can achieve and accuracy of 54.3% by always predicting the current student turn refers to one utterance back. So again, we will treat this as our baseline, and our experiments will evaluate how much improvement we can achieve with a more sophisticated approach in comparison to these simple baselines.

Table 2: Summary of features used in machine learning experiments.

Category	Features	Definition
Lexical	Utt	the string of words appearing in the current student utterance
Length	Len	the length of the current student utterance
Topic	chageOfTopic	a binary feature that indicates whether the main topic of discussion has shifted in last utterance
speaker Info	speaker	a binary feature that indicates whether the current speaker is one student or a group of students
	lastSpeaker	A binary features that indicates whether the last speaker is a student, a teacher, or a group of students
	changeOfSpeaker	a binary feature that indicates whether the student is also the speaker in the previous student turn
LSA scores	LSA1	the LSA score representing the comparison between the current student turn and the last previous turn
	LSA2	the LSA score representing the comparison between the current student turn and the turn that is two turns back
	LSA3	the LSA score representing the comparison between the current student turn and the turn that is three turns back

In our work, we built two separate classifiers to predict stuMov and respondUtt. In both cases, we first evaluated different classification algorithms by exploring combinations of different supervised learning algorithms and different sets of features on the session 1 data. Then, to test the generality of the approach we identified as best on this data, we evaluated the best performing approach on the whole corpus.

We used three standard supervised classification algorithms that have shown promising performance in previous work (Rosé et al., 2008), namely Naïve Bayes (NB), support vector machines (SMO), and decision trees (DT). For each classifier, we explored a variety of combinations of features, many of which are made available very simply through TagHelper tools. Table 2 defines the features we used and organizes them into 5 categories. Note that what is identified as the LSA score is calculated using Latent Semantic Analysis (Foltz et al., 1998), in which we calculate the semantic similarity between two utterances using LSA's latent factor approach based on patterns of co-occurrences between words. ChangeOfTopic is the only manual feature that is tagged by a human annotator. Prior work has shown that automatic topic segmentation of conversational data is feasible (Arguello & Rosé, 2006).

In our first set of experiments, we evaluated the alternative machine learning algorithms with different sets of features. The first set of features contained only the lexical features. The next set of features included that in addition to the length feature. The next set included all of those features in addition to the topic feature. The next set included all of those features in addition to the speaker information features. The next set included those plus the LSA features. The final set included all except for the lexical features.

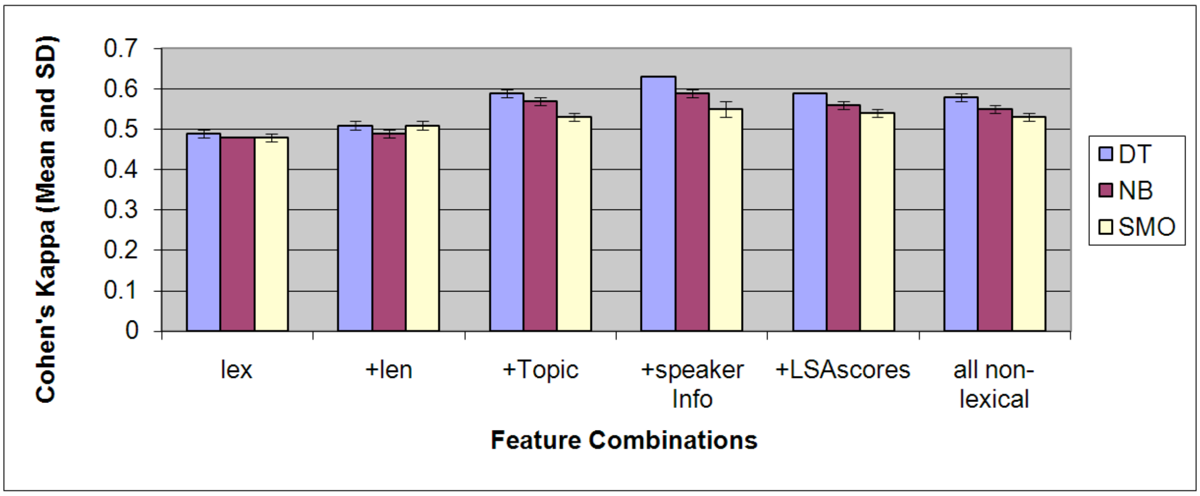

Figure 1. Classification results on predicting stuMov.

Figure 1 presents our results on predicting stuMov in the session 1 data. The bars show the Cohen's kappa and the error bars show the standard error. "+ Feature Category" stands for adding the new category of features in addition to the features that have already been used in the feature sets on its left. In the last set of feature "all non-lexical", we use all non-lexical features. We observe that the decision tree classifier trained on lex+len+changeOfTopic+speakerinfo features give us the best kappa of 0.63. Decision tree learning is also the best classifier across all feature combinations. Thus, it is reasonable to expect that this is a stable finding.

We were not satisfied with the result we were able to achieve with these combinations of features and algorithms. Thus, we conducted an error analysis in order to identify ways in which we could improve our performance. We began by looking for common confusions the best performing approach made between categories. Table 3 displays what is referred to as a confusion matrix, with allows us to see where these common confusions are occurring.

Table 3: confusion matrix of best performing classifier on predicting stuMov

	predicted NR	predicted E	predicted T
target NR	99	3	10
target E	7	19	15
target T	3	13	59

.
We can see that most of the misclassified cases are between "E" and "T". The difference between "E" and "T" is whether a student is expressing his own reasoning or relating to some previously articulated reasoning, which would have typically been contributed by a different speaker. We hypothesized that we can reduce this type of

error by taking into account more information about previous contributions where reasoning was articulated. In other words, we want to utilize the predicted student moves on previous turns. As an approximation of this type of approach, we first assign first pass categories to utterances using our three category coding scheme, and then we use these assigned categories as features in order to apply the rules found in Figure 2, which describes a heuristic post-processes stage.

Rule	Definition
1	After the teacher changing to a new topic, there shouldn't be T before E (if there is T, change it to an E)
2	After changing to a new topic, if the same student gives an NR, then teacher asks "why", the next utterance from this student is likely be E, not T
3	If a student's utterance is an E, teacher ask students to explain, then the next student utterance should be a T

Figure 2. Heuristic rules for post-processing stuMov predictions.

We implemented theses heuristic rules so that they only apply in cases where the machine learning classifiers make a prediction with low confidence. We define low confidence as confidence scores which are lower than the average confidence scores over all student utterances. Note that although the heuristic rules are manually authored, they can be executed automatically by a computer program to assign stuMov. Also, these rules only reinforce the definition of the three stuMov tags that are not captured by the classifier. Therefore, these rules are not aimed to tailor the automatically predicted stuMovs to a specific training corpus.

With the post-processing, the kappa between the predicted stuMov and the human gold standard on the first session is improved to 0.71 (SD=0.01). We further apply the best classifier and the heuristic rules on the whole corpus of all three sessions and get a kappa of 0.69 (SD=0.01), which shows our approach can be generalized to a larger corpus. This result is encouraging, but we consider it preliminary until more extensive testing on a wider variety of transcripts is conducted to ensure generalizability.

Predicting the Utterance Responded To

We employed a similar methodology in automating the analysis of which utterance a transactive utterance is responding to. We refer to this categorization as respondUtt. The target prediction is a numerical value. However, the classifiers we are using can only handle target predictions which belong to a finite set of values. Therefore, we define the class value for our prediction task to be 0, 1, 2, 4, and other. 0 stands for the current utterance, in other words, it refers to 0 utterances back, which means the current utterance is an Externalization. In contrast, values such as "1", "2", and "4" stand for Transactive moves that refer to the associated number of contributions back. If the utterance that the current utterance refers to is not 1, 2, or 4 utterances back, it is labeled as "other". We will assign a number to "other" cases in a post-processing stage using heuristic rules.

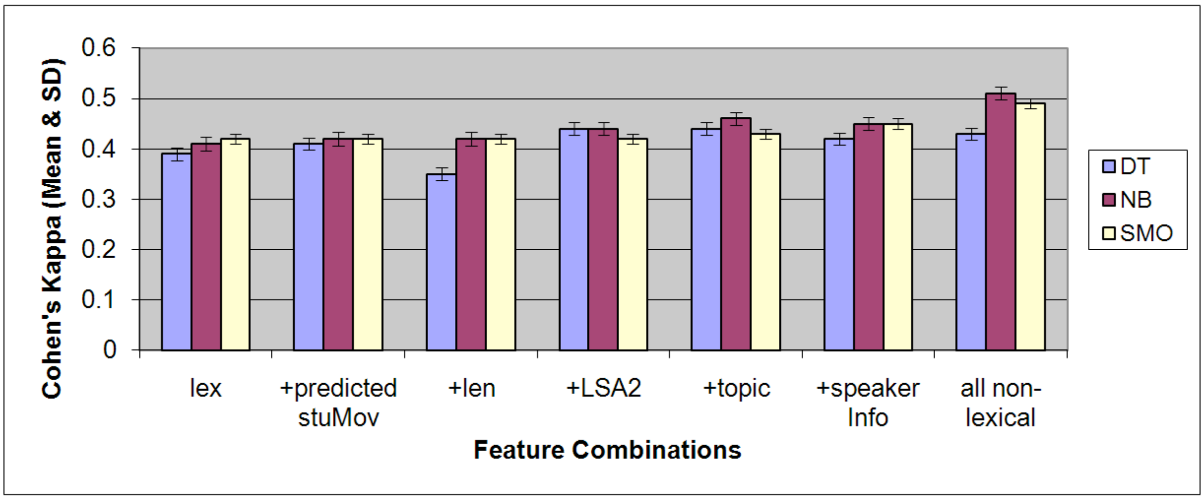

Figure 3. Classification Results on predicting respondUtt

Figure 3 shows the results on predicting respondUtt by the three classifiers in combination with the same sets of features used for predicting the stuMov. We observe that NB using all non-lexical features performs the best in this task, with a kappa of 0.51. Similarly, we add a post-processing step based on heuristic rules shown in Figure 4 using a similar methodology to what we reported in the previous section. By implementing these heuristic rules, the kappa score is improved to 0.69 (SD=0.01). Again, we apply the best performing model with the post-processing step to the whole corpus of all three sessions and get a kappa score of 0.68 (SD=0.01), which offers some verification of the generalizability of our approach.

Rule	Definition
1	A student turn with transacMov=T shouldn't point to a student turn with transacMov=NR
2	A student turn with transacMov=E, respondUtt=0
3	A student turn with transacMov=NR, respondUtt=current utterance Num-1

Figure 4. Heuristic rules for post-processing respondUtt predictions.

Discussion

Much work has already been invested in fruitful applications of automatic analysis technology in the area of computer supported collaborative learning (Wang et al., 2007; Kumar et al., 2007). The work presented in this paper points towards a new line of research applying this technology in a classroom context. One potential application of this technology could be for use in teacher professional development, by supporting instructors in reflecting on how their classroom interactions with students have proceeded and how the students are progressing in terms of striving towards articulations of transactive expressions of their reasoning. In support of the students themselves, such technology could be used to track development of a student's argumentation and articulation skills over time. Finally, eventually it may be possible for such technology to provide real time feedback to instructors or students during group discussions in order to stimulate higher levels of transactivity within the discussions. In addition to these practical applications, one can imagine that such technology also holds the potential to speed up the science of investigating the role of patterns of conversational behavior in stimulating valuable social and cognitive processes within classroom contexts. Such arguments have previously been elaborated in connection to research in computer supported collaborative learning (Rosé et al., 2008).

Conclusions and Current Directions

In this paper we have presented work to date on adapting a transactivity based analysis framework for analysis of classroom discussions and have explored the use of automatic analysis technology in this context. We demonstrated how we used an iterative development methodology to apply and extend available technology to a new analysis task, namely that of analysis of transcripts of whole group classroom discussions. Specifically, we have employed text mining techniques to analysis of transcripts from 3 classroom discussions, and have presented promising results, and positive implications both for practical applications and theory building.

Much work remains to extend this technology before real time analysis of classroom discussions will be practical. Most notably, the work presented in this paper depends upon the availability of transcripts of the classroom interaction. Thus, a considerable effort to record and automatically transcribe classroom discussions in real time using speech recognition technology still requires a substantial and targeted research effort. While advances in the field of speech recognition will undoubtedly be required in order to bring this within practical reach, prior work related to automatic assessment of group processes from speech offers hope that such an effort could be feasible (Gweon et al., 2009).

This work was supported by the Pittsburgh Science of Learning Center, NSF SBE-0836012.

References

Arguello, J. & Rosé, C. P. (2006). Topic Segmentation of Dialogue, *Proceedings of the NAACL Workshop on Analyzing Conversations in Text and Speech.*

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, 2, 202-221.

Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, 399-410.

Cakir, M., Xhafa, F., Zhou, N., & Stahl, G. (2005). Thread-based analysis of patterns of collaborative interaction in chat. *Proceedings of the 12th international conference on Artificial Intelligence in Education (AI-Ed 2005)*, Amsterdam, The Netherlands, pp. 120–127.

Chapin, S., O.Connor, C., & Anderson, N. (2003). Classroom Discussions: Using Math Talk to Help Students Learn, *Math Solutions Publications*: Sausalito, CA.

Erkens, G., & Janssen, J. (2006). Automatic coding of communication in collaboration protocols. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.) *Proceedings of the 7th International Conference of the Learning Sciences (ICLS)* (vol. 2, (pp. 1063–1064)). Mahwah, NJ: Lawrence Erlbaum Associates.

Foltz, P., Kintsch, W., & Landauer, T. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, 25, 285–308.

Gweon, G., Rosé,C. P., Albright, E., Cui, Y. (2007). Evaluating the Effect of Feedback from a CSCL Problem Solving Environment on Learning, Interaction, and Perceived Interdependence, *Proceedings of Computer Supported Collaborative Learning*

Hmelo-Silver, C. E. & Barrows, H. S. (2006). Goals and Strategies of a Problem-based Learning Facilitator. *The Interdisciplinary Journal of Problem Based Learning*, 1(1), pp 21-39.

Joshi, M. & Rose, C. P. (2007). Using transactivity in conversation summarization in educational dialog. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education.*

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education*

Kumar, R., Chaudhuri, S., Howley, I., Rosé, C. P. (2009). VMT-Basilica: An Environment for Rapid Prototyping of Collaborative Learning Environments with Dynamic Support, *Community Event Proceedings of Computer Supported Collaborative Learning*

Nathan, M., Kim, S., & Grant, T. (2009). *Instituting Change in Classroom Discourse Structure: Human and Computer-Based Motif Analysis*, Working Paper number 2009-1, Wisconsin Center for Education Research, School of Education, University of Wisconsin, Madison.

Michaels, S., O'Connor, C., & Resnick, L.B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education.*

O'Connor, M.C., Michaels, S., & Chapin, S. (2007). Small-scale experimental studies of classroom talk: Seeking local effects in discourse-intensive instruction. *Paper presented at annual meeting of the American Educational Research Association.*

Piaget, J. The *equilibrium of cognitive structures: the central problem of intellectual development*, Chicago University Press, 1985.

Resnick, L. B., Bill, V., & Lesgold, S. (1992). Developing thinking abilities in arithmetic class. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development: Implications and applications for education* (pp. 210-230). London: Routledge.

Resnick, L.B., Michaels, S., & O'Connor, C. (in press). How (well structured) talk builds the mind. In R. Sternberg & D. Preiss (Eds.), *From Genes to Context: New Discoveries about Learning from Educational Research and Their Applications.* New York: Springer.

Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F., (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning, *International Journal of Computer Supported Collaborative Learning* 3(3), pp237-271.

Schwartz, D. (1998). The productive agency that drives collaborative learning. In Dillenbourg, P. (Ed.) *Collaborative learning: Cognitive and computational approaches.* NY: Elsevier Science/Permagon.

Sionti, M., Ai, H., Rosé, C. P., and Resnick, L., A Framework for Analyzing Development of Argumentation through Classroom Discussions, in Niels Pinkwart & Bruce McClaren (Eds.) *Educational Technologies for Teaching Argumentation Skills*, Bentham Science.

Soller, A., & Lesgold, A. (2000). Modeling the Process of Collaborative Learning. Proceedings of the *International Workshop on New Technologies in Collaborative Learning*. Japan: Awaiji–Yumebutai.

Stahl, G. (2006c). *Group cognition: Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press. Retrieved from http://www.cis.drexel.edu/faculty/gerry/mit/.

Suthers, D. (2006). Technology affordances for inter-subjective meaning making: A research agenda for CSCL. *International Journal of Computer Supported Collaborative Learning*, 1, 315-337.

Teasley, S. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361-384). Berlin: Springer.

Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Wang, H. C., Rosé, C.P., Cui, Y., Chang, C. Y, Huang, C. C., Li, T. Y. (2007). Thinking Hard Together: The Long and Short of Collaborative Idea Generation for Scientific Inquiry, *Proceedings of Computer Supported Collaborative Learning*

Webb, N., Nemer, K., & Zuniga, S. (2002). Short Circuits or Superconductors? Effects of Group Composition on High Achieving Students. Science Assessment Performance, *American Educational Research Journal* 39(4), pp943-989.

Weinberger, A. & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*.

Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, second edition, Elsevier: San Francisco.