A Framework for Analyzing Development of Argumentation through Classroom Discussions

*Marietta Sionti*, Hua Ai, Carolyn Penstein Rosé, Lauren Resnick

| Language Technologies Institute-Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA,15213 |
| Psychology Department and Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA,15260 |
| Department of Linguistics-University of Athens, University Campus, Zografou, Greece |

Abstract: This chapter presents a detailed study of patterns of verbal interaction in a classroom context. In doing so it extends an important previously developed construct for analysis of productive talk for learning originating within the collaborative learning and intelligent tutoring communities, namely that of transactivity. Specifically, our focus is on argumentation and consensus building processes, which are key processes employed through language by communities in order to define themselves, maintain themselves, and evolve. We motivate the use of this construct for analysis of classroom discussions, describe our analysis framework with examples, and discuss current directions related to automatic analysis of classroom discussions using our transactivity based framework.

Keywords: Argumentation, transactivity, collaborative learning, automatic analysis technology

INTRODUCTION

In this chapter we discuss work to date on a detailed study of patterns of verbal interaction, specifically the role of social interaction through language in initiating and sustaining learning. In a broader sense, this work is also concerned with the effects of those interactions on motivation, self-attribution and commitment to a learning group that are associated with learning through social-communicative interaction, although those are not specifically in focus in this chapter. Specifically, we are investigating how human linguistic interaction works in classroom instruction and learning, and how participants in learning exchanges (both teachers and students) can best be taught productive forms of interaction. We draw from our extensive prior work related separately to classroom discourse (Resnick, Salmon, & Zeitz, 1991; Bill, Leer, Reams, & Resnick, 1992; Resnick et al., 1993; Michaels, O'Connor, & Resnick, 2007; O'Connor, Michaels, & Chapin, 2007; Michaels, O'Connor, & Resnick, 2008; Sohmer, Michaels, O'Connor, & Resnick, 2009; Resnick, Michaels, & O'Connor, in preparation) and collaborative learning (Weinberger & Fischer, 2006; Gweon, Rosé, Albright, & Cui, 2007; Joshi, M. & Rosé, 2007; Berkowitz & Gibbs, 2009).

The main contribution of this article is the expansion and explication of the analysis of what has been called “transactivity” in discourse, extended to the case of discussion in classrooms. Transactive contributions are arguments constructed in such a way as to reference, sometimes described as “operating on”, the previously expressed reasoning of self or others (Berkowitz & Gibbs, 2009). For example, consider the following dialogue excerpt from (Chapin, O'Connor, & Anderson, 2003):

S1: Well, i don't think it matters what order the numbers are in. You still get the same answer. But three times four and four times three seem like they could be talking about different things.

Teacher: Rebecca, do you agree or disagree with what Eddie is saying?

S2: Well, I agree that it doesn’t matter which number is first, because they both give you twelve. But I don't get what Eddie means about them saying different things.

Notice how the first student starts out with an attempt at expressing his reasoning about a mathematical idea. The teacher then comes in to encourage another student to attend to and address his reasoning attempt. The second student then responds, articulating not only her own reasoning,
A body of work in the collaborative learning community supports the value of this kind of transaction as a property of discussions for learning (Azmitia & Montgomery, 1993; Joshi, M. & Rosé, 2007). 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 (Bill, Leer, Reams, & Resnick, 1992; Resnick et al., 1993; Chapin, O’Connor & Anderson, 2003; Michaels, O’Connor, & Resnick, 2008; Sohmer, Michaels, O’Connor, & Resnick, 2009). 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).

We believe that the transactivity construct can usefully be applied to discussion among large groups of students in elementary and middle school classrooms, although its application is not straightforward in every case. We therefore will present our current work in developing a reliable and low-inference coding scheme to track the occurrence of transactive contributions in a teacher-led classroom discussion. Although the classroom discourse and collaborative learning/intelligent tutoring communities have proceeded mainly independently from one another, the conversational processes identified as valuable within these two communities are strongly overlapping. Our goal is to develop a framework that captures what is general across these contexts rather than being limited to any one of them. We share the intuition that the thinking/learning of an individual can be deepened, enhanced and made more robust by engaging in (linguistically mediated) interaction with other(s); more specifically, interactions that are centered on taking up the contents of an individual’s (referred to as “ego” within the transactivity framework) and another student’s (referred to as “alter”) thoughts and reasoning. In taking up alternative perspectives, piece by piece, step by step, students may challenge those contents, build on them, consider they might be integrated, and so on.

In the remainder of this chapter, we describe a theoretical framework that motivates the use of transactivity as a construct for analysis of classroom discussions. We then describe our analysis framework first at a conceptual level and then in terms of concrete coding categories illustrated with examples. We then discuss current directions related to automatic analysis of classroom discussions using our transactivity-based framework and conclude with directions for our ongoing work.

MULTIPLE THEORETICAL PERSPECTIVES

Within the field of computer supported collaborative learning, the topic of what makes group discussions productive for learning and community building has been explored with very similar findings, perhaps with subtle distinctions, and under different names such as transactivity (Bill, Leer, Reams, & Resnick, 1992; Weinberger & Fischer, 2006; Berkowitz & Gibbs, 2009) in the cognitive learning community and uptake (Suthers, 2006), group cognition (Stahl, 2006), or productive agency (Schwartz, 1998) in the socio-cultural learning community. Despite differences in orientation between the cognitive and socio-cultural learning communities, the conversational behaviors that have been identified as valuable are very similar. Schwartz and colleagues (1998) and de Lisi and Golbeck (1999) make very similar arguments for the significance of these behaviors from the Vygotskian and Piagetian theoretical frameworks respectively. The idea of transactivity originates from a Piagetian framework. However, note that when Schwartz describes from a Vygotskian framework the kind of mental scaffolding that collaborating peers offer one another, he describes it in terms of one student using words that serve as a starting place for the other student’s reasoning and construction of knowledge. This implies explicit displays of reasoning, so that the reasoning can be known by the partner and then built upon by that partner. Thus, the process sounds the same as what we describe for the production of transactive contributions. In both cases, mental models are articulated, shared, mutually examined, and possibly integrated.

Building on these common understandings, Weinberger and Fischer have developed and successfully evaluated scaffolding for collaborative learning that addresses observed weaknesses in conversational behavior related to their
In order to deepen and expand our understanding of what has been called ‘transactivity’ in the literature on collaborative dyadic interaction, we will here attempt to extend those ideas to student discourse in the context of classroom discussion. We are interested in classroom discussions because they are still the main channel of knowledge delivery in the k-12 education system. Because distance learning and online discussion forums have become more and more popular, we will extend our research to those forms of discussions later in the article.

Investigation of classroom talk in terms of the transactivity construct depends, of course, on a classroom context in which discussion takes place. Transactive classroom discussion is not the norm for US classrooms, as a number of researchers have noted (E.g., Veel, 2000; Resnick, William, Apodaca, & Rangel, in press). The traditional teaching model (emphasizing recitation, in which a teacher asks a known-answer question, the student responds, and the teacher evaluates or provides feedback) evolved (along with the lecture) as an ideal format for transmitting specific knowledge from the older generation to the younger generation. The emphasis was on the authority and authoritativeness of the instructors, with a corresponding focus on their expert knowledge. In more modern times, this ex cathedra teacher is no longer universally accepted as the ideal. Nevertheless, the characteristic pattern of the traditional teaching approach remains largely initiation-reply-evaluation (IRE), which does not afford the student the opportunity to take leadership in the public reasoning process or practice building and weighing his or her own arguments with evidence (Atkins, 2001; Core, 2003; White, 2003; Megan, 2003; Hickman, 2009).

Despite the apparent resistance to shift common practice away from the IRE recitation format to a more student reasoning-centered mode, many within the field of education have valued what is often referred to as the Socractic method, which became a popular ideal because of the idea that through Socratic directed lines of questioning students were learning the art of observation and logical induction. The goal was to lead the learner to construct his own knowledge, and the teacher would respond to a student’s answer, not with an evaluation or indication of the correct answer, but with another question or counterexample.

Dewey’s teachings (Michaels, O’Connor, & Resnick, 2007) followed those of Socrates closely in the way they emphasized dialogue and debate as fundamental principles of both democracy and education. Dewey added to this idea a theory of inquiry (Piaget, 1985; Koschmann, 2002), which can be viewed as similar to joint problem solving in collaborative learning settings.

From a separate angle, Vygotsky (1981) has argued that learning is inherently social, and that one first accomplishes in collaboration with more experienced others what one later can do on one’s own. While this does not imply that learning only occurs during social interactions, it is easy to see how his theory of learning plays out in the inter-individual interactions (e.g., dialogue) between learners, or with more experienced peers or adults. Similarly, researchers such as de Lisi and Golbeck (1999) argue that Piaget’s theory of learning applies equally to individual and collaborative learning while creating a natural place in the process for social interaction to play a key role. In their interpretation of Piaget’s theory, cooperative rather than unilateral social exchanges were valuable for countering a child’s tendencies toward either overly subjective assimilation on the one hand and overly docile imitative accommodation on the other.

The de Lisi and Golbeck interpretation of Piaget’s theory models the process through which experiences with peers can play a critical role in the development of a child’s cognitive system. A key idea that has been appropriated from this theory is that when students come together to solve a problem, bringing with them different perspectives, the interaction causes the participants to consider questions that might not have occurred to them otherwise. Through this interaction, children are said to operate on each other’s reasoning, in other words, take up and possibly transform and possibly challenge that reasoning, and in so doing they become aware of inconsistencies between their reasoning and that of their partner or even within their own model itself (Teasley, 1997). This process was termed transactive discussion after Dewey and Bentley (1949), and further formalized by Berkowitz and Gibbs (1979, 2009). A transactive discussion is defined most simply as “reasoning that operates the reasoning of another” (Berkowitz & Gibbs, 2009), although the Berkowitz and Gibbs formulation also allows for transactive contributions to operate on formerly expressed reasoning of the speaker himself.

Through engaging in reasoning and argumentation, students will improve their intellectual and thinking skills (Rqvenhscroft & McAlister, 2008). As Kuhn
and Udell (2003) experimentally demonstrated, peer dialogues generated a significant change in student ability to produce high quality argumentation in comparison to students working individually. Explicitly articulated critical reasoning and transactive discussion is what makes collaborative learning discussions valuable. When we shift to consider teacher-guided classroom discourse we will still find similar collaborative exchanges between peers, but there it will be enriched with the pedagogical lead of the teacher. The teacher is responsible for orchestrating the discussion and setting up a structure that is used to elicit reasoned participation from the students.

**TRANSACTIVE DISCOURSE IN THE CLASSROOM**

**HISTORICAL PERSPECTIVE**

Our own work has largely been located within the post-positivist, cognitively oriented research community where the primary measure of success is pre to post-test gains on academic topics, the approach to verbal protocol analysis has been primarily categorical and quantitative (Chi, 1997) and one goal has been to define patterns of conversation that can be counted and that predict pre to post-test gains. Nevertheless, we would like to stress that we greatly respect the qualitative underpinnings of a large portion of sociolinguistic work on classroom talk and collaborative discussion, especially for the rich and deep insights that work brings. However, we believe that there are general principles to be discovered within the research tradition we are working on that will be capable of distinguishing valuable instructional interactions from less valuable ones. Thus, although we do not believe that interactions are capable of causing learning in a strict sense, we believe in the capability of successful interactions to make more opportunities for learning available for children and adults. Nevertheless, we also recognize that any definition we can make precise enough to be reproducible will necessarily be a simplification.

Taking all of this into account, our short term goal is to reach a compromise, where we are able to formulate a framework that achieves a level of explanatory value with respect to pre to post-test gains sufficient to be useful for informing the design of instructional interventions while still capturing some of the richness in the data we are working with, which we also need in order to design interventions that are appropriate for their context and that won’t disrupt the learning processes we want to enhance. This chapter represents one step down that path.

Our biases in characterization of what counts as articulation of “valuable reasoning processes” grow both methodologically and conceptually from earlier work related to the self-explanation effect (Chi, 2000), which we began exploring in the context of one-on-one tutoring interactions (Rosé et al., 2001; Rosé, Moore, VanLehn, & Albritton, 2001; VanLehn et al., 2007). In study after study, the finding has been that students who self-explained more learn more (Chi, 1996; Chi et al., 2001; Chi & Roscoe, 2002).

Note that in order to compute correlations between amount of self-explanation and amount of learning, it was necessary to quantify how much self-explanation was happening. That counting process was applied to transcripts of think-aloud protocols collected as students engaged in learning activities. Note, however, that technically, self-explanation is a cognitive rather than essentially linguistic activity, although it can be observed through analysis of verbal behavior as students think aloud. As a methodological point that applies also to the discourse analysis work we are still doing, while the belief was never that self-explanation had to be audible in order to have an effect, the only self-explanation that was ever counted was what was audibly articulated, possibly with the belief that the self-explanation that was audible and therefore able to be observed at least correlated with the amount of self-explanation that was actually occurring. Similarly, we are attempting to track student reasoning processes through analysis of their discussions. We cannot know what reasoning is going on inside of students’ heads unless they articulate it. Thus, our estimation of how much reasoning is happening is almost necessarily an underestimate. However, the assumption is that there should at least be some significant correlation between the reasoning observed through the conversation and what is actually happening at a cognitive level within the individual minds of students.

We began our intellectual journey with the self-explanation construct when our own work was heavily focused on building conversational agents to act as one-on-one tutors (VanLehn et al., 2007). Our frustration at that time, however, was that while students were willing to engage in conversation with those conversational agents, they rarely gave more than one word answers, where we saw students offering much more elaborated explanations to human tutors (Rosé, Torrey, & Alevén, 2004). And
we saw the shallow interactions between the students and tutors as a severe limitation of the potential of that technology to elicit the kind of reflection we saw as valuable within the self-explanation literature. Thus, we began to explore work related to deep explanation within the collaborative learning community, with the idea that students would be more inclined to engage in deep, reflective explanation with fellow students who they saw as intellectually on par with them. The work related to the self-explanation effect connects naturally with work on elaborated explanations in collaborative learning contexts (Webb, Nemer, & Zuniga, 2004). Webb’s work provides much quantitative evidence in the form of correlational studies that elaborated explanation is associated with learning. It was our frustration with getting kids to engage in elaborated explanation in a one-on-one tutoring context that lead us to work in the collaborative learning community in the first place, where the goal was to get students to share their very different perspective with one another in order to challenge each other to think outside their own box. The goal was that in getting multiple perspectives out onto the table, students might begin to see the world from a multi-perspective point of view. We saw that students were indeed much more willing to engage in meaningful interactions with other students than with our conversational agents. And thus, we turned our attention towards the use of these conversational agents in collaborative learning settings to support the interaction between students rather than to foster an interaction between students and themselves (Wang et al., 2007; Kumar et al., 2007; Chaudhuri, Kumar, Howley, & Rosé, 2009). In this way, students had the benefit of rich interactions with their peers, but were still able to obtain correct information from the conversational agents.

LOCATING THE INSTRUCTOR

Any transcript can be coded in limitless ways. Our choice of code is driven by certain hypotheses about what kinds of peer to peer or teacher and student discourse will promote robust learning. We are seeking to make those as precise as possible, so that we can operationalize the discourse categories into a codeable form and study them systematically. In our operationalization, the status of the teacher’s moves is somewhat challenging. On the one hand, the teacher is intimately involved in the conversation, and thus it seems unnatural to separate the coding of teacher moves from those of the students. However, at the same time, the teacher’s contributions must be seen as having a special status since the teacher alone bears the responsibility for overseeing and orchestrating the interaction. Furthermore, in our work, we are investigating how teachers can be trained to behave in such a way that students benefit maximally from the classroom interactions. Thus, again, it would appear to be useful to consider the teacher’s behavior separately so that we can understand how to support it effectively. Furthermore, in separating our consideration of teacher moves and student moves, we can test hypotheses such as whether a certain sequence of teacher moves frequently lead to a certain kind of student talk or if the quantity of a particular kind of student talk is associated with better learning outcomes (e.g. pre- post-test gains).

While teachers sometimes employ moves that appear on the surface like some of the student moves we would like to see, the fact is that we are not looking for the same thing in both teacher and student discourse. The teacher may appear at times as a co-learner in modeling the types of behavior that are desired from students, but the truth is that the teacher never leaves the status of supporter, orchestrator, and primary knower. We can thus consider the teacher’s contributions to the classroom discussion as scaffolding for transactivity, or even possibly as scaffolded transactivity. Similarly Mercer talks of discussion participants scaffolding the development of each other’s reasoning through their peer interaction (Mercer, Dawes, Wegerif, & Sams, 2004). And yet, in no sense are the student’s contributions serving to scaffold the reasoning of the instructor in any real sense.

In line with all of these considerations, the authors and colleagues are collaborating through the Pittsburgh Science of Learning Center to develop two complementary coding schemes, one (discussed here) that tracks student talk, and a separate forthcoming one (Resnick, Michaels, & O’Connor, in preparation) that tracks teacher moves that scaffold transactivity development in student talk. Both teacher and student moves will then be considered together in order to cover the entire spectrum of the classroom discourse. In separating the two, we will then be in a position to study how the facilitation moves of instructors influence the occurrence of transactivity in the conversational behavior of the students.

OPERATIONALIZATION

In the context of our work on analysis of classroom discussions, work on transactivity was very attractive to us because of its emphasis both on elaborated explanation and the connection between instances of reasoning that represent different perspectives. The goal was that in making these connections, students had the opportunity to challenge their own thinking as they were faced with
the realization that it is possible to view the issue under consideration from a different view. One can think of this from the perspective of providing opportunities for cognitive restructuring to occur. Thus, one can view the goal of our analysis as an attempt to count those places in the conversation where cognitive restructuring is most likely to be triggered.

There are a variety of subtly different definitions of transactivity in the literature, however, they frequently share two aspects: namely, the requirement for reasoning to be explicitly displayed in some form, and the preference for connections to be made between the perspective of one student and that of another. Beyond that, many authors appear to classify utterances in a graded fashion, in other words, as more or less transactive, depending on two factors; the degree to which an utterance involves work on reasoning, and the degree to which an utterance involves one person operating on/thinking with the reasoning of another person. Building on this general consensus, we believe it would not be controversial to present the following student utterances as displaying a spectrum ranging from less transactive to more transactive:

| no T | S: unmodified, unsubstantiated assertion
| more T | S: The square root of 25 is 5.
| | S: externalized reasoning about one's own thoughts
| | S: First I thought -5 is smaller than -3. But now I think it's bigger, because 5 bigger than 3.
| most T | S: externalized reasoning about someone else's reasoning
| | S: I agree with him, but um I looked at the, like I found it a different way cause I thought that when you do positive, like, you still need, like you said you go to the right?

The most popular formalization of the construct of transactivity (Berkowitz & Gibbs, 1979) has 18 types of transactive moves, which characterize each child’s conversational turn, as long as it is considered an explicit reasoning display that connects with some previously articulated 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 (Berkowitz & Gibbs, 1979): elicitational, representational and operational, while a few years later (Berkowitz & Gibbs, 2009) they 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 transforming 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) (Teasley, 1997; Berkowitz & Gibbs, 2009). 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 understand from de List and Golbeck’s interpretation of Piaget (de Lisi & Golbeck, 1999) and from Azmitia’s work (Azmitia & Montgomery, 1993) that the confrontation between two points of view, also referred to as sociocognitive conflict in interaction, may cultivate a child’s reflection and ultimately learning. So it might be reasonable to hypothesize that competitive transacts might elicit more higher level reasoning than the non-competitive counterparts.

**WHAT COUNTS AS REASONING**

Our current formulation of what we’re counting as an explicit reasoning display came from work on a corpus of discussions about the design of thermodynamic cycles, which we had previously collected (Chaudhuri, Kumar, Howley, & Rosé, 2009). In this corpus, the students were pairs of sophomore mechanical engineering majors who were working with a simulation to develop an efficient power plant design. In our earlier experiences using this task in a collaborative learning setting, we had observed the conversation degenerate into a discussion about tweaking knobs and waiting for dials to move. It was therefore important to distinguish those places where the students were just doing the task without focusing on the concepts from those places where we did see evidence that the students were thinking beyond just playing with the simulation in order to achieve a high efficiency in the resulting design. One thing we saw frequently in our thermodynamics discussions, especially were statements that took the form of reasoning displays but were actually “regurgitations” of instructions the students had been given. We didn’t want to count these as
“reasoning displays” because they didn’t require the students to think for themselves beyond what they were given. Thus, we realized we needed to go beyond the surface form of the conversational contributions in our coding. 

We bring this background experience into our effort to develop a coding scheme here, however, in this chapter we focus instead on classroom discourse with much younger students. Thus, we must first decide how to adapt our earlier thinking for this new and very different context. We noticed that in our corpus of classroom discussions, the teacher begins the lesson by setting up the task that the students will engage in. So far we have treated that task setup as “what was given to students”. So if a student repeats something from that teacher presentation, we do not count that as a reasoning attempt.

A note of caution is in order here, however, since we do not want to imply a devaluing of the role of repetition within productive classroom talk. In particular, teachers frequently use repetition strategically in their orchestration of the discourse. For example, teachers may use repetition as a means for keeping an utterance vivid in order to provide a focus for the discussion. Even for students, we acknowledge that their repetition may also serve to keep the utterance vivid for themselves or their fellow students, and therefore may be a valuable step in the process. We simply do not count those repetitions as reasoning attempts.

We hesitate to assume students are always fully engaged in the attempt to reason and make their reasoning explicit. Thus, we have tried to extend the definitions from our earlier work on “what counts as reasoning for our purposes” to this data. In so doing, we have attempted to preserve Noreen Webb’s notion of “levels of explanation depth” (Webb, Nemer, & Zuniga, 2004). However, although we believe there is already a foundation of evidence that these levels of explanation depth have explanatory value with respect to test-based success criteria for learning oriented discussions, we are willing to suspend these beliefs and work towards a characterization of what counts as a reasoning display that is broader and encompasses these levels in a less “value laden” way, while still making a distinction between these levels. We do not assume either that the evidence for a reasoning display is always found within a single segment. Rather, the context can be used to illuminate what is happening within a segment.

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. At the same time, we need to distinguish attempts at reasoning from other types of contributions. In order to strike this balance, we look for evidence in the students’ articulations for attempts at reasoning displays. That will necessarily be quite subjective – especially in the case of incoherent explanations. We are continuing to work on this issue. However, we believe it is important to make it explicit that it is not a requirement of our coding that the reasoning that is displayed by students has to appear correct in order to “count”.

FORMALIZATION OF CODING CATEGORIES

In our formulation, articulation of reasoning by students is the goal, and thus we define what “counts” as a reasoning move. These are uttered by both teachers and students, but the goal is to engage students in the process of displaying their reasoning. As mentioned earlier, 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. While instructors are deeply engaged in the conversation, it is the students who are meant to benefit from the interaction. It is their articulation of reasoning that we believe is valuable for their learning. The teacher is there to support their learning, not to learn. And thus, the teacher’s status in the conversation should be treated as separate.

The teacher and students can be seen as playing reciprocal roles in that the instructor frequently scaffolds the interaction between the students, but the students never provide scaffolding to the instructor, and the instructor rarely if ever demonstrates difficulty with articulating reasoning. Our formulation of what counts as a reasoning display comes from the Weinberger and Fischer (2006) notion of what counts as an “epistemic unit”.

In that framework, 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 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 adopting it wholesale. We did this 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. In addition to a code that represents a reasoning attempt (REAS), we have additional codes for the contributions that don’t “count” as reasoning. Pure repetitions will be labeled as REPEAT. What would count as the lowest level of reasoning in Noren Webb’s framework, and thus would not count as an explicit display of reasoning in our framework, would be labeled as ASSERTION. We will also label those things that count as reasoning displays but don’t relate directly to the task that the instructor has laid out. We will label those as TANGENT. Finally, there may be blatantly off task contributions that play a purely social function or are simply not directly related to math. We will label these as SOCIAL, or the related MANAGEMENT category. Table 1 summarizes these types of contributions that do not count as attempts at explicit displays of reasoning. Utterances which belong to Not Reasoning Type 1b are blatantly off topic contributions of the SOCIAL variety and are the easiest to identify. Thus, we first check contributions for evidence of fitting into this category. Purely management oriented moves, typically uttered by the instructor, are another related category under the same type.

Table 1 Codes that refer to contributions that do not count as attempts to explicitly display reasoning

<table>
<thead>
<tr>
<th>Types</th>
<th>Label's name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>OFF TASK</td>
<td>blatantly off-task contributions</td>
</tr>
<tr>
<td></td>
<td>MANAGEMENT</td>
<td>management moves or announcements</td>
</tr>
<tr>
<td>1b</td>
<td>TANGENT</td>
<td>not related directly to the task at hand</td>
</tr>
<tr>
<td>1c</td>
<td>ASSERTION</td>
<td>plain answer or procedure reiterating what had already been articulated</td>
</tr>
<tr>
<td>1d</td>
<td>REPETITION</td>
<td></td>
</tr>
</tbody>
</table>

Both OFF TASK and MANAGEMENT moves are meant to communicate something other than specifically mathematical content. For example, the OFF TASK label includes these blatantly off-task contributions, like joking e.g.

**Teacher:** Okay, because what did you do with your two pies?

**S:** I ate them.”

The MANAGEMENT label contains all the management moves or announcements, which are usually uttered by the instructor.

**Teacher:** You don’t need to raise your hand because when you and your partner are both ready, turn and talk to each other about what you wrote, but wait till the person next to you is ready to talk.

In order to identify contributions that fit under For Not Reasoning Type 1b in Table 1, we need to consider the scope of the task that the teacher has defined. The purpose of this category is to distinguish between reasoning that addresses that task from something that might otherwise count as reasoning and might be broadly related to the topic but doesn’t directly address that task. An example of this could be where students are solving a story problem related to computing how long it takes a train to get from point A to point B, and rather than reason about the math for solving that problem, start reasoning about why a train may or may not be more efficient than a car for getting from point A to point B. We think most teachers would steer the kids away from this ancillary discussion, but maybe some would not. Another reason to represent this distinction within the coding scheme in order to preserve the ability to apply it also in the case of collaborative learning where there is no teacher present and these “off topic” conversations happen, and there is no teacher to keep them from getting off on a tangent. While this reasoning may be valuable and may be related somehow to the problem at hand, we consider it out of our scope since reasoning about these ideas may be valuable for some learning, but does not directly focuses on the concepts the teacher intended to get across with the lesson. Since the goal of our quantification of reasoning articulation is to use correlations between that quantification and pre to post-test gains, it will make the analysis less noisy to focus on those parts of the interaction that are primarily focused on the content that will be on the tests.

An utterance can be marked as TANGENT if it is not “strictly” on task or related discussion and might be valuable reasoning, might even be math, but is
not related directly to the task at hand. Similarly, the following utterance is related to the topic, which is mathematics, but it is not coherent with the rest context.

S: I know this is probably not a good time to ask this question but, um, is it possible for this to have like, like four by the power of like three or something like that?

In Not Reasoning Type 1c the statement does not give evidence of going beyond what was given (pure repetition) or going beyond a “low level” contribution according to Webb. For example it is not reasoning just to give a direct answer to a math problem with no explanation, although we acknowledge that getting the answer had to require some thinking. It is important to consider that not counting a statement as an explicit reasoning display does not mean we assume the student was not thinking. We refer to these types of contributions as with the label ASSERTION. Here is an example:

Teacher: What is the opposite of a check for ten dollars? Kevin
S: A bill for ten dollars

If the statement counts as keeping the utterance vivid without going beyond that, i.e., a pure repetition then it considered as Not Reasoning Type 1d.

In this category we mark as REPETITION every utterance that is just reiterating what had already been articulated, possibly for keeping the utterance vivid or holding the floor.

Teacher: Recall that there aren’t any negative numbers. S: There aren’t any negative numbers. Hmmm.

Utterances that have not been assigned a category by this point will typically fall under the Reasoning (REAS) category of explicitly articulating a connection between a detail related to a problem or solution and some mathematical idea. Nevertheless, it is important to note that we do not assume that the reasoning that is articulated is complete or correct. We consider as REAS the following example because it introduces a concept, in order to support the student’s objection although the reasoning displayed is not correct math.

S1: it is not a whole number
S2: yes it is
S1: cause a negative number isn’t a whole number.

Once we have identified a statement as belonging to the REAS category, we can then ask whether this is simply an externalization, which does not connect with any previously displayed reasoning, or whether it is transactive, which by definition means it connects with some previous display of reasoning. Thus, for contributions labeled with the REAS code, we go on to apply the categorization displayed in Table 2.

<table>
<thead>
<tr>
<th>REAS</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non transactive</td>
<td>Externalization (no reference to another explicit reasoning)</td>
</tr>
<tr>
<td>Transactive</td>
<td>Connection to another’s explicit reasoning</td>
</tr>
</tbody>
</table>

As Non Transactive reasoning we consider every externalization, i.e., the articulation of thoughts to the group that does not refer to previously displayed reasoning. This would include by definition any initial display of reasoning after the instructor has given the class a task to work on. Note that this does not mean that the student’s answer is lower quality. It simply means that it is an initial display of reasoning. It can then provide opportunities for students to connect to it in later contributions. Here is an example. Note that this is a good answer although it would not be coded as transactive:

Teacher: what do you think integers are and are not?
S: a fraction or it has a decimal in it cause if you look at the integers they don't have any fractions or decimals and if you look at the "not integers" they all have fractions or decimals.

As Transactive reasoning we consider contributions that display articulations of reasoning that connect with previously uttered displays of reasoning. In the following example, the teacher asks the student to provide more reasoning about his previously articulated reasoning. First we see the student’s initial contribution, then the teacher’s request, and finally the student’s self-oriented transact.

S: 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, and that [pointing to data table] just shows how much they weighed. But this shows what
they look like so you can make your own opinions.

**Teacher:** Can you say a bit more about that?

**S:** 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.

**AUTOMATIC CODING**

One purpose for formulating a coding scheme for assigning labels to types of talk moves in a formalized way is so that we can automate the analysis of classroom discussions. Much work has already been invested in fruitful applications of automatic analysis technology (Soller & Lesgold, 2000; Cakir, Xhafa, Zhou, & Stahl, 2005; Erkens & Janssen, 2006; Chaudhuri et al., 2008), especially in the area of computer supported collaborative learning (Wang et al., 2007; Kumar et al., 2007). The work presented in this chapter 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. Such an interface might present a graph displaying how distributions of teacher and student moves varied over time. Patterns that might be known from prior analyses to be associated with more or less success might be highlighted to draw the teacher’s attention to those interactions. The display could be used as an index into the recorded speech, so that instructors could then access those episodes and consider them at length. There has already been much work on development of such reporting interfaces within the computer supported collaborative learning community (Kang et al., 2008; Gweon, Kumar & Rosé, 2009), however, a parallel effort within the classroom discourse community would not be possible without first enabling automatic analysis of classroom discussions.

In support of the students themselves, such technology could be used to track development of a student’s argumentation and articulation skills over time. Again, these ideas have been popular for quite some time within the computer supported collaborative learning community (Kreijns, 2004; Soller, Mones, Jermannn, & Muehlenbrock, 2005). Within the classroom context, eventually it may be possible for such technology to provide real time feedback instructors and/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 (Chaudhuri et al., 2008).

We are beginning to explore adaptations of our earlier work on automatic analysis of typed interactions (Chaudhuri et al., 2008) and small group interactions in speech (Gweon, Kumar, & Rosé, 2009) as they can be applied to classroom discussions. Our initial work so far with this data and coding scheme has been applied to transcribed versions of classroom discussions (Ai, Sionti, Wang, and Rosé, 2010). Our results to date have been encouraging, reaching as high as .69 Kappa for distinguishing between non-reasoning, externalization, and transactive reasoning, and .68 Kappa for identifying which previous utterance a transactive utterance relates to. These initiation results were obtained using a cross-validation methodological applied to 3 class discussions with the same classroom. With this data, we cannot test the generalizability of the model beyond the specific group of students it was trained on. And we know from our prior work that groups develop a characteristic dynamic that distinguishes it from other groups, and that can lead to degradations in the performance of trained models when they are applied to different groups of kids than they were trained on. Thus, we must consider these results preliminary for now.

**CONCLUSIONS AND CURRENT DIRECTIONS**

In this paper we have presented work to date on adapting a transactivity based framework for analysis of classroom discussions, and have explored the use of automatic analysis technology in this context. We are continuing to refine our formulation of transactivity for classroom discussions as well as the work on automatic analysis.

Other technical challenges remain before this approach will be usable for automatic analysis of argumentation in a classroom context. Most notably, the work presented in this chapter depends upon the availability of reliable 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, Kumar, & Rosé, 2009).

One important direction towards improving our classification accuracy going forward is pushing our operationalization of valued student contribution types to be more precise and articulated in a way that will be easier to identify even for human analysts. In order to achieve reliability of human coding with our current operationalization, two human coders needed to go through multiple iterations of independent coding of discussion transcripts, comparison of codes, and then revision of the definitions.

In order to address this difficulty, we are seeking to improve our operationalization by drawing from insights we can obtain from work in the systemic functional linguistics community. The field of systemic functional linguistics provides a firm foundation in analyses of genres of writing (Hyland, 2000; Martin & Rosé, 2003; Martin & White, 2005), as well as face-to-face interaction (Christie, 1996), characterized in terms of the choices authors and speakers make about how to present themselves through language (Halliday, 1994). In particular, the work related to the Engagement metafunction (Martin & White, 2005), inspired by Goffman’s notion of footing (Goffman, 1979), allows us to characterize a conversational contribution in terms of the propositional content communicated, the source of that content, the author/speaker’s attitude towards that content, the assumed attitude of listeners towards that content, as well as the speaker’s alignment or misalignment with the listeners and/or the source of the content. We mentioned earlier that the Berkowitz and Gibbs formulation of transactivity has 18 categories, which distinguish at a conceptual level different ways one explicit reasoning display may make a connection with another explicit reasoning display. What we believe we can draw from systemic functional linguistics is a language level vocabulary for making these types of connections explicit, in terms of how they are encoded in language. Beyond this, we can view transactive contributions in terms of their social implications as well as the cognitive ones, which have been the focus of our early work described in this chapter.

Beyond aiding us in improving our ability to identify transactive moves in conversations as human annotators, we believe that what we can learn from systemic functional linguistics will also help us improve our ability to use text mining and machine learning technology to automatically identify these types of moves in transcripts of conversations. Feature space design, i.e., the extraction of features that are predictive of the target category classification, is a critical part of machine learning. This is an especially difficult challenge in the field of textual classification, where an arbitrary number of features of varying complexity can be extracted from documents as a preprocessing step. A challenge for researchers has consistently been to balance expressiveness of features with the size of the corresponding feature space, due to issues with data sparsity that arise as feature spaces grow larger. However, if even moderately more complex features are added to the feature space, such as the class of features that describe grammatical relationships that might hold between pairs of words within a sentence, the size of the feature space drastically expands, which frequently reduces rather than improves the effectiveness of the feature space representation (Joshi & Rosé, 2009). At the same time, moderately more complex features such as these do very little to improve the representational power of the feature space. What is needed is dramatically more complex features. However, if even expanding to moderately more complex features is on the edge of computational feasibility, then expanding to this greater level of sophistication might seem to be beyond reach.

Nevertheless, we believe we are well positioned to move in this direction. We will draw upon recent successes utilizing genetic programming (Koza, 1990; Koza, 1992; Koza, 1994) in problems with similar feature space representation challenges outside of language processing tasks (E.g., Krawiec, 2002; Smith, & Bull, 2005) as well as within language processing (Rosé, 1999; Rosé, 2000; Hirsch, Hirsch, & Saeedi, 2007). We have already begun to achieve promising results through a technique for constructing complex features from simpler features, and adding these more complex features into a combined feature space which can then be utilized by more advanced machine learning classifiers. In our recent work applying this technique to a sentiment analysis problem (Mayfield & Rosé, 2010), we show significant improvement in classification accuracy over a simpler, state-of-the-art baseline feature space, with a small and constant increase in feature space size. Our plan is to pursue this approach, using insights from systemic functional linguistics in the construction of novel features.

ACKNOWLEDGEMENTS
REFERENCES


Sohmer, R., Michaels, S., O’Connor, M.C., & Resnick, L.B. (2009). Guided construction of knowledge in the classroom: The Troika of well structured talk, tasks, and tools. In B. Schwarz & T. Dreyfus (Eds.), Advances in Learning and Instruction, pp. 105-129
Vygotsky, L. (1981). The genesis of higher mental functions