Tracking Changes in Classroom Discourse Structure Using Human Pattern Identification and Computer Based Motif Analysis

Mitchell J. Nathan, Suyeon Kim and Timothy S. Grant

University of Wisconsin-Madison

1025 West Johnson St., Madison WI 53706-1796

Phone: 608-263-0563 (Nathan)

Abstract

We compared the structure of discussions in a middle school mathematics classroom before (Year 1) and after (Year 2) teacher participation in professional development activities aimed at enhancing students' participation and the co-construction of mathematical ideas. Changes in the role of the teacher and student are accompanied by identifiable changes in discourse structure, not just content. In particular, while traditional Initiation-Response-Evaluation (IRE) patterns continue throughout, the de-centering of the teacher leads to a reduction in IRE occurrences and increases in student led demonstrations of math ideas that receive peer-generated evaluations and elaborations. Analyses conducted by human coders were corroborated by computer-based motif analyses that identified flexible patterns using probabilistic, data-mining methods, while also predicting novel discourse structures.

When people talk—at a café, the workplace, in classrooms—they tend to adopt regular patterns of organization. What forces influence the organizational structures of classroom discourse? In this paper we compare the structure of classroom discussions before and after teacher participation in professional development activities that were aimed at enhancing students' classroom participation and the co-construction of mathematical ideas. We show that changes in the classroom are accompanied by identifiable changes in discourse structure, not just changes in content. In particular, the de-centering of the teacher's mathematical authority leads to a reduction in traditional Initiation-Response-Evaluation (IRE) patterns, and increases in student led demonstrations of math ideas (IDE patterns) that receive peer-generated evaluations and elaborations.

Theoretical Framework

Monologic discourse focuses on response to an authority in a highly expected manner, often as a way to show adherence with the canon (Bakhtin, 1986; Hakkarainen & Paavola, in press). Wells & Arauz (2006) point out that monologic instruction is sometimes necessary to propagate knowledge from previous generations, but it is not sufficient to serve as the sole vehicle for instruction**.**

> Not only do children not always understand what they are told and so need to engage in clarifying dialogue to reach the desired intersubjectivity, but frequently they also have alternative perspectives on a topic that need to be brought into the arena of communication and explored in more symmetric dialogue in which there is reciprocity in the roles of speaker and listener, and equally, an attempt by each to understand the perspective of the other. (p. 387)

Dialogic discourse, in contrast, derives from a *participatory* view of learning (Sfard, 1998, 2008) that frames knowledge as distributed and culturally bound, emphasizing the socially mediated nature of distributed, personalized knowledge generation. It is also a powerful way to promote student engagement and higher order reasoning (Nathan & Kim, 2007), long-term retention, and transfer of concepts to new contexts. In a dialogic exchange the teacher frequently elicits student involvement with open-ended questions, prompts students to share multiple perspectives, and invites further ideas through follow-up questions/utterances (Nassaji & Wells, 2000; Wells & Arauz, 2006).

To transform monologic instruction into dialogic interactions within classroom, the locus of authority of knowledge must be de-centered, and students granted permission to initiate discussion, have protracted turns-at-talk where they explore their mathematical ideas, and legitimately provide evaluations of the accuracy and appropriateness of a peer's contributions.

Everyday Talk and Institutional Talk

The everyday talk of casual conversation and domestic affairs enjoys regular organizational constraints (Sacks, Schegloff & Jefferson, 1974). Institutional talk, as found in court proceedings, medical interactions, the workplace and classrooms, differs from everyday talk (e.g., Drew & Heritage, 1992) because those in positions of authority have ethical and professional obligations constraining the prompts they offer and their responses. For example, in a medical interaction, the doctor may have a moral imperative to ask questions of a lay patient, who, in turn, is obliged to respond truthfully to meet the intended aims of the interaction (ten Have, 1999). As another example, during an interview, the turns of asking and answering questions are pre-allocated based on the status of the interlocutors; interviewers have interactional authority and responsibility to ask questions, whereas the role of the interviewee is constrained to answering them.

In contrast to the two-part sequences stereotypical of conversations and question-answer exchanges, pedagogical interactions commonly exhibit three-part sequences. This triadic organization emerges from the overlapping role of two contiguous adjacency pairs. The first part--typically a question or initiation from a teacher--calls for a match in the form of a student reply to the complete pair. Because of professional obligations facing the teacher, this reply, while serving as the second part of the triad, also is the *first part* of a new adjacency pair, and so warrants its own conditionally relevant response, such as an authoritative evaluation of the accuracy or appropriateness of the student response.

Patterns such as initiation-response-evaluation (IRE) sequences are pervasive during pedagogical exchanges (Greenleaf & Freedman, 1993; Mehan, 1979; Sinclair & Coulthard, 1975), and often dominate educational discourse (e.g., Lemke, 1990; Wells, 1993). In a typical IRE pattern, the teacher initiates the exchange (an I-event) by asking (most often) a closed, “known-answer” display question. This elicits a direct reaction (R-event) from a student, whose response is then evaluated (E-event) by the teacher, often in a way that terminates the interaction. While these typically arise in classrooms, Sefi (1988) showed that the talk during a home-visit from a health practitioner to a new mother is similarly organized in three-part sequences (question-answer-comment) because of the didactic nature of the interactions.

Evaluative and Non-Evaluative Exchanges

Cullen (2002) distinguishes between the evaluative and the discoursal roles of the third element of the interaction. In the evaluative case (IRE), teacher feedback is given to directly accept or reject the student response, laying judgment within a context of a power imbalance, that often fails to elicit further participation (cf. Nystrand, 1997). Thus, we observe a monologic exchange (Bakhtin, 1986; Lottman, 1988) that may be useful in delivering information and assessing knowledge but does little to evoke multiple perspectives, support the construction of common ground, and provide for chaining of prior talk (Lemke, 1990; Mercer, 1992; Nathan, Eilam & Kim, 2007; Nunan, 1987; Nystrand, 1997; Thornbury, 1996; Wells & Arauz, 2006; Wood, 1992).

When the evaluation phase is replaced with a non-evaluative follow-up question (IRF), it provides more impetus to perpetuate the discourse (Wells, 1993; Wells & Arauz, 2006). In IRF exchanges, the follow-up movement, or *F-movement* (Sinclair & Coulthard, 1975; Cullen, 2002), invites multiple perspectives and typically leads to further contributions from students, thus perpetuating their involvement and increasing both their engagement (Mercer, 1995) and their participation in the kinds of open-ended exchanges that can promote higher-order reasoning (Fernyhough, 1996). For methodological purposes, several investigations (Cazden, 2001; Nathan et al., 2007; Wells & Arauz, 2006) combine IRE and IRF organization into a single category that emphasizes the authority-directed three-part sequence structure.

Student Agency in Classroom Discourse

These findings notwithstanding, Nunan (1987) and others (e.g., Lemke, 1990; Nystrand, 1997; Thornbury, 1996; Wood, 1992) are critical of the conditions that encourage IRF and IRE exchanges because students have limited opportunities to verbalize their own ideas. There is also mounting evidence that it fails to impart a dialogic exchange that fosters engagement and conceptual understanding (Cullen, 2002; Nystrand, 1997; Wells, 1999; Wells & Arauz, 2006).

With educational reform objectives, it is becoming common to see students initiating discursive sequences and serving as the primary agent for enacting the evaluation and follow-up

(F-movement) events (Engle & Conant, 2002; Lampert, 1990; Stipek, et al., 1998). These interactions may invoke protracted demonstrations of students' ideas that are far more elaborate and personalized than typical R-events that follow known-answer questions. Sequences built around these demonstrations, or D-events, can dominate highly participatory classrooms, and show frequent chaining. For example, Nathan and colleagues (2007) documented IDE sequences in 77.8% of the exchanges of a whole-class collaborative problem-solving session, with chaining from prior to subsequent solutions in 81% of the exchanges. The demonstrations elicited student-directed responses the vast majority of the time (93%) that tended to integrate evaluative and elaborative (non-evaluative) statements (E events) that invited further participation of other speakers with alternative perspectives. Nathan and colleagues suggested that the precipitating influences of the IDE structure might be found in the open-ended nature of the initiations posed (usually by the teacher), and the established norms of interaction and the creation of a respectful and secure classroom environment, which enabled students to dominate the discussion and serve as principle agents for evaluation and for directing the subsequent discourse. These types of classroom exchanges constitute *productive discourse*; that is, "forms of social exchange which provide participants with an avenue to construct and build upon mathematically correct conceptions through their interactions with other class members" (Nathan & Knuth, 2003, p. 204). Those in the classroom built on each other's ideas, even when they did not agree, and they listened to, reflected on, and were genuinely interested in, each other's solution proposals (Rommetveit, 1989).

Computer-Directed Discourse Analysis

There are a number of computer-based and computer-assisted systems for scoring and analyzing free-form discourse. Many of the most effective systems receive streams of text rather than spoken language (Litman, Rose, Forbes-Riley, VanLehn, Bhembe, & Silliman, 2006). A good example is the TagHelper system developed by Rose and colleagues (Rosé, Wang, Cuie, Arguello, Fischer, Weinberger, & Stegmann 2008). The TagHelper tool set operates with raw text and employs both feature based rules and statistical machine learning to determine the categories of discourse events that are exhibited in a stream of online text generated during collaborative work. Automated processing of spoken language presents many additional challenges, especially in naturally occurring settings where the quality of the speech signal cannot be carefully controlled and there are multiple, and even overlapping, speakers. Computer based text analysis systems also perform to varying degrees of accuracy depending on the restrictions placed on the discourse. When investigators tightly constrain the topics and forms of speaker interactions, the automated analysis can achieve relatively high levels of performance (Pennebaker & Francis 1996). However, when the constraints on the discussions are eliminated performance of the system declines considerably.

The main motivation for computer based discourse analysis is efficiency of processing data. The more aspects that can be off-loaded to the computer, and the less time spent by humans, the greater the efficiency. Human beings pay central roles in two aspects of the automated process. First, humans are involved in pre-processing of the data, so that the incoming stream of text or speech falls within the standards of the computer algorithm. "Hand coding" of data is inevitably part of the training set of even the most automated system. Second, there is a need for trained human coders to check some portion of the coded data to establish reliability.

Rose and colleagues (2008) point out that because the underlying processes employed by people and by computer programs are fundamentally different, even when reasonable (quantitative) levels of agreement are established between humans and machine, the nature of disagreements are not the same as the disagreements that arise among human coders. In reporting on their experiments with the automated text categorizations using the TagHelper system, the algorithms make types of errors that would be highly unusual for people.

It is important to note that all computer-based systems of discourse analysis are the product of both human and computer processes, and in this sense they all lie along a continuum of computer assisted text processing. The Motif analysis method that we describe in this paper occupies a middle ground of automation and level of algorithmic sophistication with regard to language processing, with significant human input at both the input and output stages.

Empirical Hypotheses

Our expectation is that changes in the locus of authority in the classroom will be reflected as identifiable changes in the event sequences in the discourse structure. Specifically, we anticipate that the event sequences in the Year 1 lessons will show a strong presence of IRE/F patterns, with teacher-directed questions eliciting short responses from students that are then evaluated by the teacher. The nature of the classroom changed significantly in Year 2, and students served a more central role in participating in and directing the discourse. We expect this form of exchange to be more supportive of IDE event sequences than that evident in the lessons from Year 1. As a secondary research focus, we expect the computer-based motif analysis to corroborate the human coding of the data. We also leave open at this early stage in our work the possibility that the motif algorithm may reveal recurrent patterns within the classroom data that have heretofore not been identified. However, we invite this new source of analysis as a means to further advance the research on classroom discourse.

Method

Participants and Setting

We observed a sixth-grade mathematics class operating within a public middle school (grades 6 to 8) over a 2-year period. The students exhibited a wide range of mathematical performance in each of the two years (standardized tests from 5th to the 99th percentile). The teacher's training was in elementary education, and while she was the senior math teacher at the time she also taught sixth grade science, language arts, and French. She was nominated by her principal to participate in this study of classroom discourse. She expressed a strong professional urge to develop her classroom discourse techniques, so that she could implement the kinds of instruction called for in reform documents such as the NCTM (1989, 2000) *Principles and Standards*.

For this study, we arbitrarily selected two lessons each from Year 1 and Year 2 from a larger corpus, on the criteria that they each contained an extended, whole classroom discussion about mathematics that was consistent with the teacher's planned curriculum, and therefore constituted ordinary instructional time.

Professional Development Intervention

Over a three-week period during the intervening summer, the teacher participated in conversations about reform instruction, her professional and personal goals as a teacher, and video-prompted reviews of her Year 1 classroom teaching, as well as that of others (e.g., Ball, 2003; Victoria Zack, personal communication). Our sessions included readings from math teachers and educational research about learning, curriculum and participatory forms of instruction, how to achieve classroom norms conducive to student participation and student-directed learning, such as active listening, ways of presenting one's ideas to the room, how to model giving constructive feedback, and how to elicit and use students' multiple solution methods to facilitate mathematical participation and learning.

Coding and Pattern Identification

Each video was transcribed, segmented into analytic units, then further subdivided into coded events such as initiation (I-events), response, demonstration, evaluation, elaboration, and F-movements, consistent with prior work reported by others (Lemke, 1990; Mehan, 1979; Nathan et al., 2007; Sinclair & Coulthard, 1975). In addition, event codes took into account agency (student, teacher or conjoint), and other qualities that were relevant to the investigation (e.g., short or protracted responses, open or closed questions, etc.). The complete set of event codes, along with examples, is shown in Table 1.

Pattern identification using human coders. The role of previously documented patterns such as IRE, IRF and IDE provided theoretical guidance for the pattern-finding process, though coders were open to finding other patterns as well. Table 2 shows an example sequence of event codes taken from Year 1, Lesson 1. Pattern finding consisted of iterative examination of the string of codes, the transcript, and the video using Transana (Fassnacht & Woods, 2005), a computer program for qualitative and quantitative digital video and audio analysis.

Pattern identification using motif analysis. The challenge confronting human coders is that patterns that may exhibit some variation in their surface structure may actually represent sufficiently similar underlying discourse structures as to warrant inclusion. *Motifs* are essentially flexible patterns that allow for systematic identification of these underlying structures. Motif analysis uses random walks based on probabilities to find likely patterns and potential starting sites within the coded sequences (Cadez et al., 2003; Grant, 2007; Keles et al., 2002). For example, a portion of the code stream from Lesson 1 from Year 1 of our corpus is shown in Figure 1. The algorithm begins by segmenting the original data stream (Figure 1a) of temporally ordered codes from the transcript (Figure 1b; we chose segment lengths of ten codes), then establishing a random starting point (Figure 1c). Each segment is assigned a motif at random (e.g., Dg-EE-Ti; Figure 1c) and a site (4^{th} position) at which the motif starts. Starting motifs (Figure 1d) are defined based on the segments that are assigned a common motif. (Note in the example that the algorithm is considering a DEI pattern, which, when part of a chain, may be an IDE motif.)

Figure 2 illustrates the random walk of the motif algorithm as it identifies patterns. A selected segment (Figure 2, Step 1) is identified by the particular motif and site at which the motif begins. The possible steps for the random walk are determined by three pieces of information: The sites at which the motifs start, the set of motifs present at that site, and the probabilities that define each of the motifs. This particular study uses Gibbs sampling (ref) to first select a set of probable starting sites and estimate which motifs are likely to begin at each site. First, a segment is selected (Step 1). Based on the current definitions of motifs (Step 2), new motifs and starting positions are selected at random with probabilities defined by how likely each motif is to start at each position (Step 3). Then, the motifs are redefined based on the new assignments of motif and starting position (Step 4). This process is repeated thousands of times (we chose to use 50,000 trials before reporting each result). Over the course of the random walk, the definitions of motifs and the assignment of motifs will converge to within a region of probable (though not necessarily optimal) solutions.

Trade-offs between different approaches to pattern identification. Both the human and computer based approaches draw from the same set of codes, and therefore operate with common inputs. Human pattern finding is driven not only by perceptual processes, but also by hermeneutic considerations where meaning making and interpretation shape the outcomes. Because of this the process is inherently subjective. People are also subject to severe constraints on attention and working memory load, as well as fatigue. Though detailed records were not kept, the human directed pattern finding processes took over 3 months even after transcripts were furnished.

Under motif analysis, some biases are removed from the pattern identification process, particularly in the exploratory mode, because the search and construction of motif definitions is purely data-driven rather than goal directed. The algorithms draw on vast memory stores and can therefore simultaneously maintain an arbitrary number of provisional patterns until the system settles on a set of dominant sequences to hold over for future consideration. In this regard, the computer algorithms will find whatever flexible patterns are present in the coded data strings (Table 1). Biases in the initial coding that are fed to the pattern finding algorithm are still possible, but become decoupled from the pattern identification process. Thus, given our coding system and alphabet, the algorithm has no *a priori* preferences favoring identification of IRE, IDE, XYZ, or any other event sequences. The process is also highly reliable, being the product of thousands of independent trials before the findings are reported. The pattern finding cycle is also considerably faster, taking on the order of one minute for each data set (approximately 100,000 times faster than the human coders).

Limitations to this approach are that the process is a meaningless activity to the program and can result in identifying patterns that are difficult to interpret in discourse terms. One way to compensate for this is to review each dominant pattern. Motif analysis locates the sites in the data stream for each dominant pattern that is found. This allows the research team to inspect the section of video for each finding to ensure it fits the criteria for the types of classroom interactions, rather than being an artifact of the algorithm. In this way, the process is best described as a *computer-assisted* approach, rather than one that it completely automated. We return to this point in the Discussion section.

Results and Conclusions

We analyzed a total of four classroom lessons from the same teacher, two from Year 1 and two from Year 2. Lessons were selected so that they included whole classroom instruction. The data for each class episode were segmented into events as described in Table 1 above. The inter-rater reliability for the coding showed 94.5% agreement (*kappa* = .94).

Pattern identification by human coders. Example excerpts of the different code sequences are shown in the Appendix. As Table 3 shows, across the two Year 1 lessons the IRE (n = 25) and IRF (n = 15) patterns as identified by the human coders were dominant. IDE sequences were extremely rare in Year 1 (n = 1). The remaining event sequences did not fit a pre-ordained category.

In Year 2, the pattern identification showed a marked change in the discourse structure. Consistent with the predictions about the influences of a shift in the teacher's approach to discourse in the classroom, IDE pattern use increased (n = 18). IRE and IRF patterns, while still prevalent, occupied a smaller portion of the discourse (IRE n = 9; IRF n = 11).

Following scholars such as Wells (1993), we combined IRE and IRF patterns since both had teacher initiated discussion, short student responses, and teacher-led evaluations or follow-ups. A chi-square analysis on frequency of IRE/F vs. IDE patterns in Years 1 and 2 was significant, Chi-Square (1) = 33.5, $p < .01$. Consistent with the stated hypotheses, there was a greater presence of IRE/F patterns in the first year than in the second year, which saw an increased presence of IDE.

Computer-generated motifs: Exploratory analyses. The human identification of the event sequences suggests that there are identifiable patterns within the discourse. However, as is evident from a cursory inspection (Table 2), there are many complexities to the data, including event sequence variants, insertions, and partial sequences, which make identifying patterns cognitively demanding, fraught with subjectivity, and potentially low reliability. We used the motif analysis in the exploratory, or data-driven, mode for motifs of length 3, so it was particularly tuned to find IRE, IRF, IDE sequences). We also explored a motif analysis with length 4, to provide context for these patterns (Keles, 2002).

For Year1 Lesson 1, the dominant motif that emerged with a window of length 3 was an IRE pattern (including IRF sequences), evident in 20 of the 23 segments (Table 4a). When a window of length 4 was used, the dominant motif that emerged in 21 out of 26 segments was E-IRE (Table 4b). This reflects one of the ways the larger pattern window shows the context of these patterns, and the algorithm deals with the IRE pattern chaining throughout the lesson, with E-events from a prior IRE sequence serving as the trigger for a subsequent IRE sequence. Year 1 Lesson 2 showed the same dominant motifs for length 3 (Table 4a) and length 4 (Table 4b), but the algorithm also predicted they would occur less frequently.

The second year lessons, following the professional development, showed some departure from the IRE pattern, with the emergence of more open-ended initiation prompts, followed by student demonstrations of their own mathematical ideas. In the length 3 motif analysis (Table 4a), the dominant motif showed a mix of the IRE, IRI and IDE sequences. IRE continues to be a presence in the Year 2 data. However, IRE now shares its impact with other event patterns. The sudden prominence of the IDE pattern is in keeping with our initial predictions concerning changes in the discourse structure following changes in the classroom climate.

The emergence of the new IRI (see Year 2 Lesson 1, where I can substitute equally for E in the ending position) sequence shows some of the power of the motif algorithm as a bottom-up means for identifying patterns that may otherwise pass unnoticed by human coders. In this pattern the question poser (typically the teacher) reiterates basically the same question following a response. Note that it is distinguished from IRF sequences in that the utterance following the response does not build on the response in any way (no F-movement; Sinclair & Coulthard, 1975). Rather it is a reflection that the respondent must have misheard or misunderstood the initial inquiry.

As Table 4a shows, there were also strongly competing alternate motifs, specifically DEI and IR(I or E) (not shown). The algorithm appears to have found a point of multiple equilibria, where no single motif received enough of an advantage to become dominant, and segments just moved equally between them. This reflects the Bayesian nature of the motif analysis, where the system seeks to “settle” in a (set of) stable states that reflect the most prominent matches. When different patterns compete equally as dominant patterns in the data, the system literally bounces between the different states, as, for example, in Rubin’s vase-face optical illusion. This reflects some of the complex dynamics of classroom discourse. Note that some of the dominant and alternate motifs (e.g., DEI, EIR) can be reordered to make an IRE or IDE pattern, which essentially reveals the influences on pattern identification of chaining of recurrent patterns, where an E or D from a prior triad occupies the first position of a subsequent IRE or IDE sequence.

The length 4 motif for Year 2 Lesson 1 (Table 4b) showed that multiple codes were substitutable in the first and third sites. Even so, this flexible pattern is strong and was predicted to appear in 13 of the 20 segments (65%). The pattern also showed a leading evaluation and short-answer response events that may precede (and trigger) IRE and IDE patterns. When R occupies the first position, this again shows the new IRI pattern, which, when chaining, can be part of an event sequence such as RIRI…, and so on.

Year 2 Lesson 2 clearly shows that the discourse was at times organized around student-led demonstrations in response to initiations, along with a reduction in display (closed) questions posed by the teacher (Table 4a). The dominant motif of length 3 found in 8 of the 16 segments is characterized as an IDE pattern, and so takes up about half of all the class exchanges. The longer, length 4 motif (Table 4b), shows again the leading context of an evaluation, and the substitutability of codes, this time in the third site, where R- and D-events appear to play a similar role in the discourse structure, and reveal the co-dominance of IRE and IDE motifs.

Computer-generated motifs: Confirmatory analyses. Three patterns generated by the human analysis and the exploratory motif analysis (length 3) were selected for specific investigation by means of a confirmatory motif analysis: a traditional IRE pattern, the IDE pattern, and the novel IRI sequence. For this analysis the algorithm is operating in a top-down mode, rather than the bottom-up mode used in the exploratory analysis. Table 5a shows the frequencies with which each of these motifs (as defined in Table 5b) occur in the data. As predicted, the IRE pattern was the major organizational structure in the Year 1 lessons, even taking into account the longer length of the first lesson. We also found presence of a new form of

IDE pattern in Year 1, which was initiated by a closed question, much like a traditional IRE. We label this IDE* to indicate it is a hybrid sequence (Table 5a), in that it shows a protracted demonstration in response to a closed, teacher-initiated question, followed by a teacher evaluation or elaboration (Table 5b). In essence, it begins and ends like an IRE, with a student demonstration in the middle.

IRE also played an important role in Year 2. In further support of the central hypotheses, the proper IDE pattern, with an open-ended question and either a student or teacher evaluation or elaboration following a student demonstration, was clearly evident in Year 2 ($n = 19$) though it was negligible in Year 1 ($n = 2$), as shown by both the human and computer coding. We were also able to confirm the reliability of the novel IRI pattern across all the lessons as one that emerged from the confirmatory analysis. As noted, this is potentially an important discovery made by the motif analysis since it was not identified by the human coders and has not been previously documented in the literature on classroom discourse.

The probabilistic structure of motifs allowed us to quantify how unlikely it is that these patterns occur in their reported frequencies by chance alone (Table 6). The less likely they are to occur, the more support we have that these do actually reflect patterns in the data. For our cut-off, we use occurrences of less than 1 in 10,000 ($p < .0001$). IRE patterns in all four lessons were significant, as were IDE patterns in Year 2. Once again, the IRI pattern was confirmed, though it was only statistically reliable Year 2 Lesson 1 using our very conservative threshold. This provides further support for the hypothesized changes in discourse structure with the changes in the classroom environment.

Comparison between human and computer-based pattern identification. As a final analysis, we provide a comparison of the sensitivity of human coders and the motif algorithm for identifying the major event sequences, IRE/F and IDE. For simplification, we combined IRE, IRF and IRI into a single IRE category, and combined IDE* and IDE into a single IDE category (Table 7). Figure 3 makes several points visually apparent. First, the motif analysis tends to be more sensitive, especially for the IRE patterns. This is most likely due to the probabilistic nature of the motifs that allow for code substitution in the definitions of the flexible patterns. Second, both human and computer methods show, qualitatively, strikingly similar patterns over time, providing a form of corroboration of the general patterns within the data. IRE patterns are in far greater numbers initially and maintain a presence in the later lessons; while IDE patterns show an increased presence in the lessons from Year 2, with usage rising to the IRE levels by the final lesson.

The juxtaposition of human and computer-based findings shows corroboration of these methods, but also lends support to this new data-mining method for automatically finding new patterns in large sets of noisy data. Two discoveries are of particular note. First, flexible patterns showed substitutable forms, as when R and D events were of comparable dominance during the early Year 2 lesson. This suggests that different events (i.e., event codes) may, from a linguistics perspective, be processed similarly by classroom participants. Second, an entirely new motif, the IRI sequence, went undetected by the human pattern-finding process, but emerged from the data and passed through the stringent criteria of the confirmatory analysis. The motif algorithm also found an interesting variant of the IDE pattern (which we labeled IDE*), which presents a hybrid between IRE and IDE sequences.

In summary, we found that changes in the climate of the classroom that invite greater student participation in the mathematical interactions can lead to identifiable changes in the structure of the discourse. In Year 2, the class enacted far more IDE sequences than Year 1, though IRE still remained an important construct throughout the data. In the final section we consider the nature of the influences on classroom discourse structure, and reflect on the methods for documenting these influences using both human and computer based methods.

Discussion

Using both human and computer-based methods we were able to show that structural aspects of the discourse exhibit change when the teacher makes an intentional shift in the locus of authority, and encourages a more student-directed learning environment. As predicted, Year 1 lessons in our sample tended to follow the IRE/F structure that has typically been found when teachers employ didactic instruction (Cazden, 2001; Mehan, 1979; Sinclair & Coulthard, 1975). While IRE patterns maintained a presence in Year 2, we also saw growth in IDE patterns, reflecting the greater role of students to provide protracted reports on their mathematical thinking and to evaluate and elaborate on the thinking exhibited by their peers (Nathan et al., 2007). Accompanying this shift, and perhaps instrumental to it, the teacher seeded discussions with more frequent uses of open-ended questions.

The shifts in discourse structure in the second year of this data set align with what Engle & Conant (2002) characterize as the *principles of productive disciplinary engagement of students.* They argue that the instructor needs to both encourage students and provide adequate resources for them to be stakeholders in the intellectual problems at hand. This means students as well as the teacher need to initiate questions, rather than just providing responses to teacher's queries; learners need feel they have the authority to pursue these inquiries; students need to be held accountable to both their peers and the disciplinary norms, where "accountability does not require acceptance of others' views, but instead responsiveness to them" (p. 405); and students need sufficient time and resources to pursue a problem in depth. According to Engle & Conant, when given the opportunities to experience productive engagement students develop greater reliance on evidence-based argumentation, they reevaluate their beliefs, and refine their own positions by evaluating the ideas promoted by others. In many respects, we have observed some of these developments.

In addition to documenting these changes in discourse structure, we showed how both human and computer-based methods could be used to document them. The two sets of findings showed remarkable corroboration (Figure 3). Underlying this is an important synergy that was established between the two analytic methods. The motif analysis, while computer based, is really a *computer-assisted* process, because it was necessary to first code the data manually, and then to move between the numerical output of the program and the human interpretation of the transcript that the motif sites referred to. It is only with these hermeneutic influences that the motifs gain any meaning whatsoever, and lend support to the analyses of the classroom interactions.

However, the motif analysis is not simply an automated version of the human process. When run as a purely bottom-up method it exposed a novel event sequence, the IRI pattern, which went undetected by the human coders. This sequence is in evidence when a primary speaker feels the need to reiterate the original question rather than use a further question to expand or challenge a response (an F-movement). As such, one would expect it to be more prevalent early in the process of developing one's open-ended questioning style, as we found in the Year 2, Lesson 1 data. The motif analysis also identified a novel hybrid IDE pattern (IDE*) that incorporated the initiation and evaluation events most commonly found in traditional IRE patterns, though these occurred in conjunction with student demonstrations of mathematical knowledge. Future work on the transitions to student-directed discourse would be useful in further understanding the nature of the IDE* sequence and the role it may play as teachers develop new discourse repertoires in their classrooms.

One of the other novel findings is the identification of multiple equilibria in the discourse, where two or more event sequences establish prominence and vie for dominance. The presence of these equilibria reminds us of the complex dynamics of group discussions, and the tremendous flexibility that agents exercise during participation. Further work in this area may shed light on how complex, socially mediated learning settings operate on a systemic level, and may help to identify when and how such interactions develop into productive forms of interaction (Engle & Conant, 2002), as well as how they move toward convergence (Kapur, Voiklis, Kinzer, & Black, 2006).

This investigation highlights two critical aspects of the study of classroom discourse. First, the structure of events among interlocutors is both indicative of and responsive to climate changes in the classroom. Second, as discourse analysis methods and theories of the nature of group discussions evolve, the types of exchanges that will be identified will also change. As these findings make their way to the practitioner literature and to teacher education and professional development programs, this will, in turn, affect the classroom experiences. This underscores the dynamic interplay between the nature of the phenomenon under investigation and the methods of analysis employed.

References

Bakhtin, M. M. (1986) Speech Genres and Other Late Essays. Trans. by Vern W. McGee. Austin, Tx: University of Texas Press.

Ball, 2003)

Cadez, I., Heckerman, D., Meek, C., Smyth, P., & White, S. (2003). Model-based clustering and visualization of navigation patterns on a web site. *Data Mining and Knowledge Discovery*, 7, 399-424.

Cazden, C. B. (2001). *Classroom Discourse*. Portsmouth, NH: Heinemann, 2001

Cullen, R. (2002). Supportive teacher talk: The importance of the F- move. *ELT Journal*, *56* (2), 117-127.

Drew, P., & Heritage, J. (1992). *Talk at work: Interaction in institutional settings*. Cambridge, England: Cambridge University Press.

Engle, R. A. & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399-483.

Fassnacht, C., & Woods, D. (2005). Transana v2.0x [Computer software]. Available from http://www.transana.org

Fernyhough, C. (1996). The Dialogic mind: A dialogic approach to the higher mental functions. *New Ideas in Psychology*, *14*, 47-62.

Grant, T. S. (2007). An Implementation of Motif Data Mining on Protocol Reports for Contextual Patterns. Unpublished paper. University of Wisconsin-Madison, School of Education.

Greenleaf, C., & Freedman, S. W. (1993). Linking classroom discourse and classroom content: Following the trail of intellectual work in a writing lesson. *Discourse Processes*, *16* (4), 465–505.

Hakkarainen, K. & Paavola, S. (in press). Toward trialogical approach to learning. To appear in Schwarz, B., Hershkowitz, R., & Dreyfus, T. (Eds.) Guided Construction of Knowledge in Classrooms. Elsevier Advances in Learning and Instruction Book Series.

Kapur, M., Voiklis, J., Kinzer, C., & Black J. (2006, June). Insights into the emergence of convergence in group discussions. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings* *of the International Conference of the Learning Sciences* (pp. 300-306). Mahwah, NJ: Erlbaum.

Keles, S., van der Laan, M.J. & Eisen, M.B. (2002). Identification of regulatory elements using a feature selection method. *Bioinformatics*, 18, 1167-1175.

Lampert, M. (1990a). Connecting inventions with conventions. In L. P. Steffe & T. Wood (Eds.), *Transforming children's mathematics education* (pp. 253–265). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lemke, J. L. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.

Litman, D. J., Rose, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D. and Silliman, S. (2006) Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education, 16*, 145-170.

Lotman, Y. M. (1988). Text within a text. *Soviet Psychology, 26*(3), 32–51.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Mercer, N. (1995). *The guided construction of knowledge: Talk among teachers and learners*. Clevedon: Multilingual Matters.

Nathan, M. J., Eilam, B., & Kim, S. (2007). To disagree, we must also agree: How intersubjectivity structures and perpetuates discourse in a mathematics classroom. *Journal of the Learning Sciences*, *16*(4), 525-565.

Nathan, M. J. & Kim, S. (2007). Regulation of teacher elicitations and the impact on student participation and cognition. WCER Working Paper Series. Wisconsin Center for Educational Research: Madison, WI.

National Council of Teachers of Mathematics. (1991). *Curriculum and evaluation standards for school mathematics*. Reston, Virginia.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Nunan, D. (1987). Communicative language teaching: Making it work. *ELT Journal, 42* (1), 136-145.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. Martin Nystrand with Adam Gamoran, Robert Kachur, and Catherine Prendergast. New York: Teachers College Press.

Pennebaker, J.W. & Francis, M.E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion, 10*, 601-626.

Rosé, C. P., Wang, Y-C., Cui, Y., Arguello, J., Fischer, F., Weinberger, A., & Stegmann, K. (2007). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer-Supported Collaborative Learning*.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turns-taking in conversation. *Language, 50*, 696–735.

Sefi, S. (1988). Health visitors talking to mothers. *Health Visitor 61*(1), 7-10.

Sfard, A. (2008). *Thinking as Communicating: Human Development, the Growth of Discourses, and Mathematizing* (Learning in Doing: Social, Cognitive and Computational Perspectives) Cambridge University Press: London.

Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.

Stipek, D., Salmon, J. M., Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education, 29,* 465–488.

Thornbury, S. (1996). Teachers research teacher talk. *ELT Journal, 50* (4), 279-289.

ten Have, P. (1999). *Doing Conversation Analysis: A Practical Guide*. Thousand Oaks: Sage.

Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education, 5*, 1–38.

Wells, G. & Arauz, R. M. (2006). Dialogue in the classroom, *Journal of the Learning Sciences, 15*, 379-428.

Wood, D. (1992). Teaching talk. In K. Norman (Ed.), *Thinking voices: The work of the national oracy project* (pp. 203–214). London: Hodder and Stoughton for the National Curriculum Council.

Appendix

Example Excerpts of the IRE and IDE code sequences

Excerpt 1. IRE event sequence (from Year 2-2).

T: [Name] what do you think?

S1: Um, I know what the, uh, even number is for that. How to bound it into a whole number.

T: Good.

Excerpt 2. IDE event sequence initiated by Teacher.

T: (name), what do you think?

S : ((pointing to the fraction on the board)) Well, I have, well we, I didn't do it that way but um, I think the five would mean um, how, how many hours it would take to do half the wall. That sounds what it's like saying, because you're taking the ten and like dividing it in half, it's like you're diving the wall in half.

S: If you, if you have a full wall, it's (inaudible) ten hours.

Excerpt 3. IDE event sequence initiated by Student.

S: Now, I don't see why you (inaudible) because, two by two.

S: ((Pointing to the fraction on the board)) There are two walls right here and they'd still be painting one wall, so you need to divide it by two.

S: No, but, that is if there were (inaudible). if that's minus (inaudible). If you're saying, okay, I'm just going to paint (inaudible) but they're saying he's going to paint until (inaudible) done then (inaudible).

Tables & Figures

Table 1. The event codes used, along with examples from the transcripts.

Code	Description	Criteria and example
Ti	Teacher's display question	→ T: So what does one represent? One hour? S: Yeah
Si	Student's display question	→S: Then, what else is blue? S2: Five.
Bi	Teacher's & Student's display question	→S: Do you want us to draw one of the other dots too? →T: Sure, can you find another? S: Um, two-fourths.
TI	Teacher's Open question	→ T: would you be willing to show us why you got five? And I'll be interested to see the reasoning.
SI	Student's open question	→S7: Who would like to speak? →Amy: Now can I speak?
BI	Teacher's & Student's open question	→ S: What's an improper fraction? →T: What does that mean, improper?
RR	Student's short, direct verbal response	T: what color? → S: Yellow.
DD	Student's demonstration with drawing	S: Jones' one hour and combined them together ((coloring one column in a table on the board)), like that one and that right there ((Drawing a new vertical line in one column in the table)).
DG	Student's demonstration with gestures	T: So, I just want her to talk about her technique. →S: So he did that much in an hour, and she did that much in an hour ((pointing to one column with an index finger and pointing to another column at the bottom table on the board))
Dg	Student's gesture –only demonstration	T: Come up and point for me. →S: ((Pointing one point on the graph on the OHP)
DW	Student's demonstration with writing	S :And then, I just write times seven over seven ((writing a formula "7/7" on the board))
TF	Teacher's F-movement	S: I'm going to add these two. → T: Why are you going to add these them?
SF	Student's F-movement	S6: Blue. →S7: How did you get that?
BF	Teacher's & Student's F-movement	S2 : I'm thinking that they split the wall in half. →S3 : But why would it, (inaudible) any higher if (inaudible) hours. →T: What do you think about her question, Jane?

Tf	Teacher's subsequent F-movement	S: Well, like, it's an odd number so you can't really have.. →T (TF): Which is an odd number? S: Seven. →T (Tf): Oh, seven's an odd number?
TE	Teacher's valenced evaluation	S: One half. → T: Good.
SE	Student's valenced evaluation	T: What's the point of this? S5: To see the number…. →S6: No, no, no.
NE	Teacher's neutral evaluation	S: I added them. →T: Okay.
Te	Teacher's elaboration	S: If you, if you have a full wall, it's (inaudible) ten hours →T: He's, he's saying, He is saying that if you add three and seven to get ten, that's really two walls. That's Miss, Miss Jones doing a whole all and Mr. King doing a whole wall.
Se	Student's elaboration	S1: It's either two hours or a half, or one half hour. →S2: Or four hours
EE	Teacher's evaluation & valenced elaboration	S: Twenty one over, twenty one over twenty one. →T: Right. Twenty one over twenty one would be exactly one. So it's really close to one.
ee	Student's valenced evaluation & elaboration	S1: It was yellow. →S2: One, two, three, four, five, yellow. →S3: No. No. It's because if alright, so if the tenth one was yellow.....
BE	Teacher's & Student's valenced evaluation and/or elaboration	T: If you used ten as your .. as your numerator? John says twenty, yes? S: Yeah. →T: So it's really close to a half. Isn't it like really close. → S : If you double it, then twenty twenty ones.

Table 2. Example section from Year 1, Lesson 1, of the stream of codes obtained from the 4 lesson transcripts used for both human and computer-based pattern finding. For the human coder, the matches are: Yellow highlight = IRE (defined with this sequence of substitutable events: Ti-RR-TE/NE/EE/Te), Blue = IRF (TI/Ti-RR-TF/BF/Tf), and Green = IDE (TI/SI/BI-DD/DG/Dg/DW-TE/Te/SE/Se/EE/ee/BE/ Be).

Ti	Ti	RR	Ti	Ti
RR	RR	Te	RR	RR
TF	Te	Bi	TE	Te
		RR		
Dg	TI	Te	Bi	Ti
EE	DG	Ti	RR	RR
Ti	Te	RR8	TF	TE
DD		TI	RR	
EE	Ti	RR	Te	Ti
	RR	NE		RR
Ti	NE	TI	Ti	TE
RR		RR	RR	
EE	Ti	NE	TF	Bi
	Dg			RR
Ti	TE	TI	RR	EE
RR	Ti	RR	TE	
Te	Dg	TF		Ti
	Te		Ti	RR
Ti	TF	RR	RR	TE
Dg	RR		TE	
TF	EE	Ti		Ti
RR	Ti	RR	Ti	RR
Tf	DW	NE	RR	TE
RR	TE		TF	
		Ti		Ti
Ti	Ti	RR	TF	RR
RR	RR	NE	RR	TE
TF	TF		NE	Ti
			Ti	RR
			RR	NE

Table 3. Class time for the lessons in Years 1 and 2 and frequency of IDE, IRE and IRF patterns identified by human coders.

	Class Time	IDE	IRE	IRF
Y1-1	18:50	1	20	11
Y1-2	15:21	0	5	4
Y2-1	20:33	7	3	7
Y2-2	20:42	11	6	4
Total		19	34	26

Table 4. Results from the exploratory motif analysis for (a) a motif window of length 3, and (b) and motif window of length 4.

	Frequency of Each Motif					
Session	Dominant Motif (Length 3)	Dominant Motif	Alternate Motif	Alternate Motif	No Motif	Total Segments
Y1-1	IRE	20	1	0	2	23
Y1-2	IRE	8	1	1	1	11
Y2-1	IR(IE) or DEI	9	4	1	3	17
Y2-2	IDE	8	4	2	2	16

	Frequency of Each Motif					
Session	Dominant Motif (Length 4)	Dominant Motif	Alternate Motif	Alternate Motif	No Motif	Total Segments
Y1-1	EIRE	21	2	1	2	26
Y1-2	EIRE	7	2	0	3	12
Y2-1	(RE) I (RD) E	13	2	2	3	20
Y2-2	EI (RD) E	14	4	0	1	19

Table 5. Results of the confirmatory motif analysis. (a) Frequency of patterns found (with motif window of length 3). (b) Codes from Table 1 used in the definition of each motif.

Confirmatory Results (Length 3 Window)				
Motif	Y1-1	Y1-2	Y2-1	Y2-2
IRE	32	16	10	11
IRI	17	6	15	7
IDE Total	6	1	7	11
IDE* (Ti = Closed Question)	4	1	0	0
IDE (TI, SI or BI = Open Question)	2	0	7	11
Total Number of Codes	186	84	138	126

Motif IRE	1st Site	2nd Site	3rd Site
	Ti TI TF	RR	Ee EE NE Te TE
Motif IDE*	1st Site	2nd Site	3rd Site
	Ti	DD Dg DG DW	Ee EE NE Te TE
Motif IDE	1st Site	2nd Site	3rd Site
	TI SI BI	DD Dg DG DW	Ee EE Te TE Se SE Be BE NE

Table 6. Probabilities of observed occurrences based on frequencies of each code.

Confirmatory Results (Length 3 Window)				
	Y1-1	Y1-2	Y2-1	Y2-2
Motif IRE/F	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Motif IDE	0.25	1.0000	< 0.0001	< 0.0001
Motif IRI	0.0003	0.0725	0.0001	0.15
Number of Codes	186	84	138	126

Table 7. Comparison between number of event sequences identified by human coders and computer motif algorithm, combining like sequences (see Figure 3).

Lesson	Human IRE/F (IRE + IRI + IRF)	Computer IRE/F (IRE + IRI + IRF)	Human IDE (IDE* + IDE)	Computer IDE (IDE* + IDE)
Y1-1	31	49	1	6
Y1-2	9	22	0	1
Y2-1	10	25	7	7
Y2-2	10	18	11	11

a. Original session

Ti RR TF Dg EE Ti DD EE Ti RR EE Ti RR Te Ti Dg TF RR Tf . . . Ti RR

b. Cut string into segments of length L = 10 (with

Ti RR TF Dg EE Ti DD EE Ti RR
Ti RR EE Ti RR Te Ti Dg TF RR
TF RR TF RR Tf RR Ti RR TF Ti
et

c. Randomly assign starting site and motif selection

Ti RR TF Dg EE Ti DD EE Ti RR   Motif 1 at Site 4
Ti RR EE Ti RR Te Ti Dg TF RR   Motif 2 at Site 6
TF RR TF RR Tf RR Ti RR TF Ti   Motif 1 at Site 2

d. Starting motifs defined

If Motif 1 is made of Dg EE Ti and RR TF RR

Code	1st site	2nd site	3rd site
Dg	0.50	0	0
EE	0	0.50	0
Ti	0	0	0.50
RR	0.50	0	0.50
TF	0	0.50	0

It is assumed for this example that only the first and third segments are associated with motif 1 at the start of the algorithm. If motif 1 is present starting at sites 4 and 2 respectively, the motif would initially be defined as in the table above with each cells recoding the probability of that site (1st, 2nd, or 3rd) receiving the appropriate code (Dg, EE, Ti, RR, or TF).

Iterative steps

STEP 1. Select a segment.

Ti RR TF Dg EE Ti DD EE Ti RR

STEP 2. Based on defined motif's determine probabilities for all combinations

Motif 1 Site 1: Ti RR TF Probability = 0.01
Motif 1 Site 2: RR TF Dg Probabilitly = 0.01
…
Motif 3 Site 8 EE Ti RR Probability = 0.03

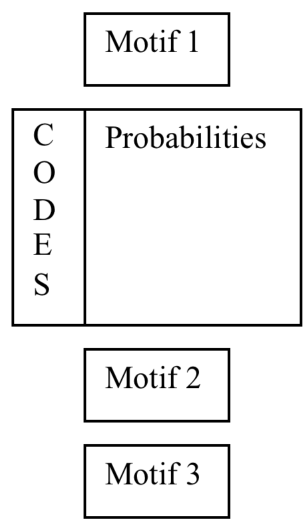
STEP 3. New site and motif selected at random based on probabilities. So Motif 3 at Site 8 is 3 times as likely as Motif 1 at Site 1. For demonstration, select Motif 1 starting at Site 6.

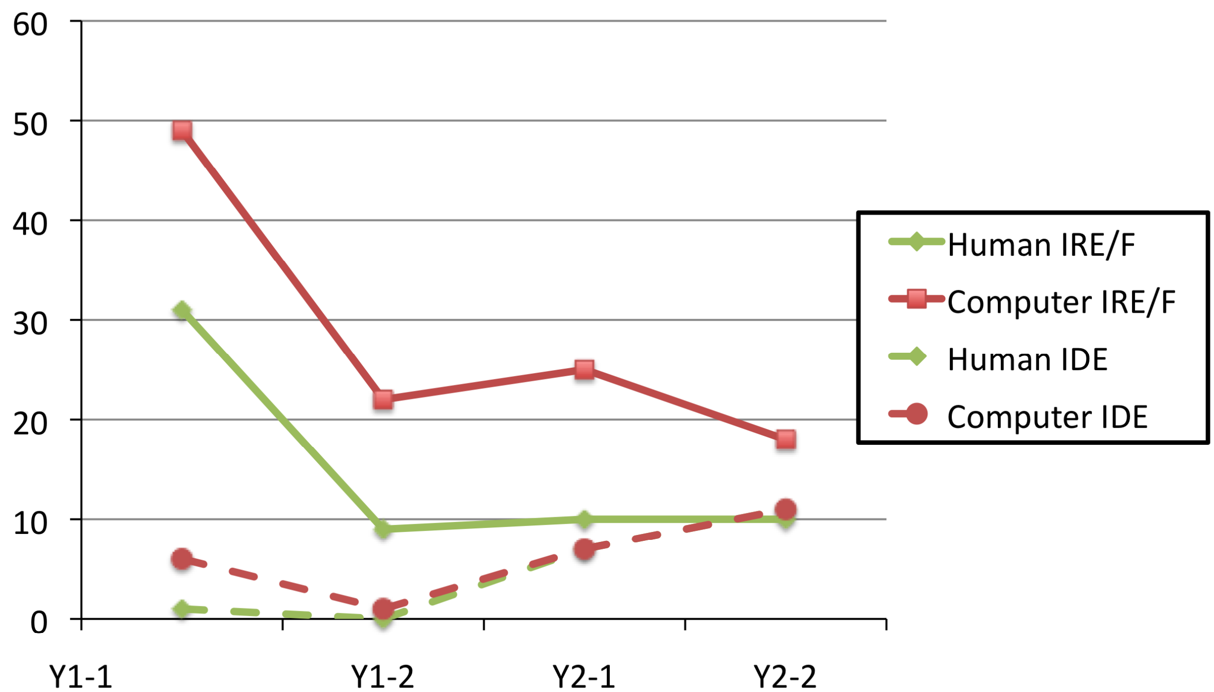
Ti RR TF Dg EE Ti DD EE Ti RR Motif 1 at Site 6

STEP 4. Redefine motifs and go to STEP 1. (Repeat 20,000 times)

Over time the definition of each motif settles, and sites where each motif occurs are identified. Output is a list of the assignment of motif and starting site for each segment and the definition of the motifs.

Segment	Motif	Start
1	3	4
2	2	6
Last	3	3




60
50
40
30
20
10
0
Y1-1
Y1-2
Y2-1
Y2-2
Human IRE/F
Computer IRE/F
Human IDE
Computer IDE