Chapter 23

Combining Coding and Conversation Analysis of VMT Chats

Alan Zemel, Fatos Xhafa & Murat Çakir

ARZ27@Drexel.edu, FXhafa@uoc.edu, MPC48@Drexel.edu

Abstract: This chapter considers the relationship between statistical analysis of coding based on theoretical schemes and conversation analysis of VMT participants' structuring of their chats. It describes how a statistical test on a hypothesis regarding collaboration in VMT showed an unexpected result, whose understanding required the use of qualitative methods. The phenomenon behind the puzzling result was identified using conversation analysis. The chapter explores an approach to coding based on analysis of how sequences of discussion of different topics are defined interactionally by chat participants as accomplishments of their postings. A form of "mixed methods" is proposed using codes for the different sequences and displaying the ordering of these longer sequences of interaction or compiling statistics of these codes.

Keywords: Statistical analysis, conversation analysis, expository participation, explanatory participation, long sequences, probability transition tables

The analysis of the use of software by groups is particularly problematic. Most methods of human-computer interaction were developed for single-user systems and are not applicable to computer mediation of group interaction. A common approach to analyzing the use of groupware is to compare statistical measures of usage across conditions or cases. However, this can be criticized for not investigating and taking into account qualitative differences that may be crucial to understanding the quantitative differences. While there is a widespread feeling that fields like CSCL and CSCW need to take a multidisciplinary approach incorporating a variety of analytic methods, it is difficult to see how quantitative and qualitative approaches built on fundamentally incompatible theoretical foundations can be synthesized. This chapter reports a case in which a quantitative finding motivated a qualitative analysis to explain the significance of the statistical results. This experience suggested to us a novel approach to combining the two: using qualitative analysis to derive the coding scheme for quantitative analysis.

In the VMT Project, we have investigated online problem solving from a variety of analytical and methodological perspectives. In our first year, we developed a coding scheme and applied it to logs of online chats among actors participating in math problem solving (Chapter 22). The coded logs were intended to provide a basis for quantitative analysis of the chats. While we were still investigating the coding approach, we also became interested in conversation analytic methods as a way of describing the procedures participants use to make sense of their ongoing activity. Conversation analysis (CA) and statistical analysis (SA) are uneasy partners in the analytic enterprise. These two orientations to analysis derive from very different perspectives on the role of the analyst and the kinds of assumptions that can be made with respect to the data and its interpretation.

In *statistical analysis*, hypotheses are put forward and tested. Coding schemes are devised that are designed to facilitate the testing of these hypotheses and statistical methods are applied to the coded data. In this approach, it is the analyst's perspective that is privileged. The analyst:

- Proposes the hypotheses,
- Produces the coding scheme to capture the relevant data from an experiment designed specifically to allow for testing of the hypothesis, and
- Assesses and interprets the statistical results (Mason, Gunst & Hess, 2003).

Statistical analysis of data gathered from online collaborative learning experiments plays a central role in many CSCL studies (e.g., Avouris & Margaritis, 2002; Daradoumis, Martínez & Xhafa, 2004; Dillenbourg et al., 1996; Strijbos, 2004). A whole range of statistical methods—from descriptive statistics to multilevel and other sophisticated methods—have been used to analyze the underlying features (variables) of the collaborative activity that takes place in a small group.

Conversation analysis, on the other hand, is an analytical methodology that attempts to describe the actions of participants in terms of the relevancies demonstrated by participants through their interaction (Pomerantz & Fehr, 1991; Psathas, 1995; ten Have, 1999). Actions are analyzed as situated within a stream of ongoing action and as sequentially organized. Furthermore, conversation analysts presume that actors design their action to fit the particular circumstances in which they are accomplished—and which they thereby reproduce, extend and help constitute.

The differences between SA and CA are consequential. For statistical analysts, validity and reliability are significant concerns (see Chapter 22). However, these are not concerns for conversation analysts because CA has a different view of the nature of the data. For SA, the analysis of data is to be conducted through what statisticians consider to be objective procedures that control for subjectivity and bias. In contrast, CA takes the data as already meaningful in the eyes of the participants and therefore open to being understood by analysts (who share membership in the social and linguistic cultures of the participants). Conversation analysts are concerned with providing adequate descriptions of the sense-making procedures used by participants as they interact. Where statistical analysts want to discover frequently observed regularities in interactions, conversation analysts are concerned with how specific actions were made relevant by prior actions and how a current action makes relevant subsequent actions over the course of a particular sequence of actions. For conversation analysts, it is sufficient that the participants in a particular interaction treat their ongoing actions as sensible. The conversation analyst's task is to describe these sequences of actions as sense-making procedures. SA assumes a causal model of behavior and tries to confirm predictive statistical patterns, whereas CA looks for non-deterministic social methods that people use as interacting agents.

While these two types of analysis—statistical and conversational—may seem incompatible, it turns out there are circumstances in which they can be mutually informative (Heritage & Roth, 1995). In this chapter, we describe a situation in which a puzzling statistical result was made intelligible by conversation analytic investigation. This is a novel approach to analyze the organization of the interaction in collaborative math problem-solving activities in online chats. Indeed, existing approaches in the literature treat quantitative and qualitative methods separately, often relegating the qualitative to pre-scientific exploration or post-scientific speculation. Our results show the strength of using a combined approach. Specifically, by using a quantitative approach, we detected an unexpected result in a hypothesis test. This made further investigation necessary. The qualitative method of CA enabled us to identify the phenomenon that produced the unexpected result in the SA hypothesis test.

The Statistical Analysis

We took the six chats discussed in Chapter 20 (see Table 20-1, reproduced as Table 23-1). In each chat, a group of 3 to 5 students in grades 6 to 11 collaborate online synchronously to solve math problems that require reflection and discussion using AOL's Instant Messenger software. We coded each chat using the scheme discussed in Chapter 20 and analyzed in greater detail in Chapter 22. The coding scheme includes nine distinct dimensions, each of which is designed to capture a certain type of information from a different perspective. The coding scheme was synthesized from research in CSCL, adapted through trial with VMT data (as described in Chapter 22). Two dimensions coded the threading (see Chapter 20) in order to unpack the response structure—which might otherwise lead to confusion in analyzing the flow of interaction (see Chapter 21). The other dimensions were intended to capture the content of the session. This chapter considers only the content-based dimensions: conversation, problem solving, social reference, math moves and system support.

Table 23-1. Description of the coded chat logs.

PoW-wow Session #	Facilitator	Members	Number of Postings	PoW Name	Announced Before?
1	MUR	PIN, GOR, REA, MCP	334	Finding CE	No
2a	GER	AVR, PIN, SUP, OFF	724	Equilateral Triangle Areas	No
2b	MUR	MCP, AH3, REA	204	Equilateral Triangle Areas	No
9	POW	EEF, AME, AZN, LIF, FIR	715	Making Triangles	Yes
10	MFP	AME, FIR, MCP	582	The Perimeter of an Octagon	Yes
18	MFP	AME, KOH, KIL, ROB	488	A Tangent Square and Circle	Yes

Recall that the sample of six chats is made up of three in which the math problem was announced at the beginning of the session, whereas in the rest the problem was posted on the Math Forum's web site in advance. It should be noted, however, that announcing the math problem in advance doesn't necessarily mean that the participants of the chat already solved the problem in advance.

To see what we could learn from statistical analysis after putting in a major effort in developing the coding scheme and coding six full chats, we looked for statistical differences between the chats by students who knew the problem before working together ("known") versus the chats by students who did not ("not").

Our first objective was to test whether there is any significant effect of the "known/not" criterion on the sample of the six chats ("PoW-wows"). To this end, we started by computing, through descriptive statistics, the distribution of frequencies in different dimensions (conversation, social reference, problem solving, math move and system support) for the six PoW-wows; we used Means and ANOVA to test the existence of significant differences due to the known/not criterion. The study showed that there was no such effect, at the usual confidence level of 95% (in fact, significance in differences, that is significant pairs, were not noticed even at a 90% confidence level). The fact that there is no clear effect of the criteria known/not prompts us to conclude that the classification of the sample of PoW-wows into groups according to the known/not criterion is not relevant. We could also observe this by computing the box-plot representation of the variables under study (see Figure 23-1).

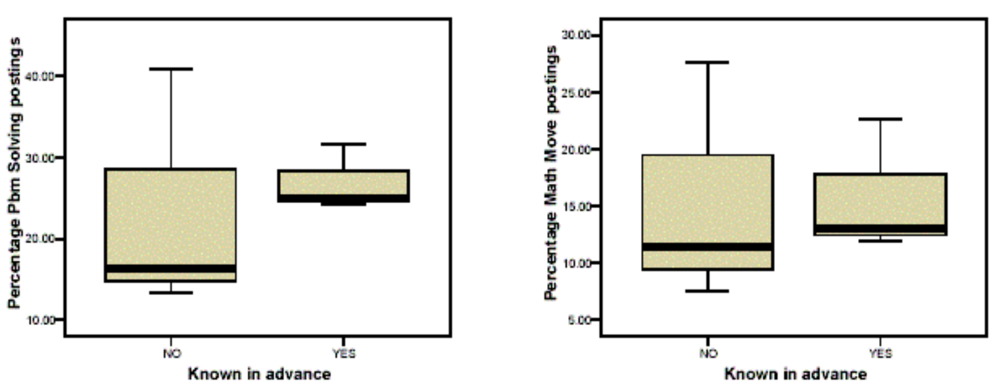

Figure 23-1: Box-plots of problem-solving and math-move dimensions.

Given the above finding, we refined the statistical analysis by looking at the correlation between vectors of values of the six PoW-wows—we continued to group by ‘known in advance’/‘not known in advance’ just for visual effect. By computing similarities between the PoW-wows we could see which PoW-wows are similar to each other and which are different from each other. We computed the correlations (Pearson correlations) in the proximity matrix shown in Table 23-2.

Table 23-2. Pearson correlation of vector values of 6 PoW-wows.

	Pow1: Not	Pow2a: Not	Pow2b: Not	Pow9: Known	Pow10: Known	Pow18: Known
Pow1: Not	1.000	0.756	-0.452	0.567	0.108	-0.197
Pow2a: Not	0.756	1.000	-0.219	0.912	0.603	0.067
Pow2b: Not	-0.452	-0.219	1.000	0.202	0.620	0.956
Pow9: Known	0.567	0.912	0.202	1.000	0.867	0.470
Pow10: Known	0.108	0.603	0.620	0.867	1.000	0.791
Pow18: Known	-0.197	0.067	0.956	0.470	0.791	1.000

From Table 23-2 we observe the following:

- Pow2b (Not) is negatively correlated to the other two PoW-wows of the Not group (Pow1 and Pow2a) and positively correlated to the PoW-wows of the Known group (Pow9, Pow10, Pow18). Moreover, significant correlations of Pow2b (Not) with Pow10 (Known) and Pow18 (Known) are observed and a non-significant correlation with Pow9 (Known).
- There is a significant positive correlation of Pow9 with Pow1 and Pow2a of the Not group. In pair-wise terms, Pow9 is more correlated to the PoW-wows of the Not group than to the PoW-wows of its own Known group.
- There are some pairs of PoW-wows positively and strongly correlated, namely (Pow2a, Pow9) and (Pow2b, Pow18) which suggest taking a closer study of the possible common features of these PoW-wows.

The previous observations on the correlations between PoW-wows from different groups not only support the claim that there is no significant effect of the known/not criterion, but also shed light on the reason why these two groups are not really separated. Indeed, the negative correlation of Pow2b with the PoW-wows of the Not group shows that its place is not in the Not group. Even more, its positive correlation with the PoW-wows of the Known group indicates that this PoW-wow is better grouped with the PoW-wows of the Known group.

In our next step, we decided to exclude the system-support dimension from the analysis; indeed, this dimension is less relevant in the context of the interaction analysis, and could have introduced some noise in the analysis. We ran the statistical computations again by re-computing the correlations in the proximity matrix shown in Table 23-3.

Table 23-3. Pearson correlations with system support excluded.

	Pow1: Not	Pow2a: Not	Pow2b: Not	Pow9: Known	Pow10: Known	Pow18: Known
Pow1: Not	1.000	0.999	-0.427	0.868	0.376	-0.145
Pow2a: Not	0.999	1.000	-0.396	0.884	0.407	-0.112
Pow2b: Not	-0.427	-0.396	1.000	0.080	0.678	0.957
Pow9: Known	0.868	0.884	0.080	1.000	0.787	0.366
Pow10: Known	0.376	0.407	0.678	0.787	1.000	0.862
Pow18: Known	-0.145	-0.112	0.957	0.366	0.862	1.000

By excluding the system-support dimension, we observe a clear effect on the correlations, namely:

- On the one hand, an increased negative correlation of Pow2b (Not) with the other PoW-wows of its group (Pow1 and Pow2a) is now observed. Notice also that the correlation between Pow1 and Pow2a is almost a perfect correlation. On the other hand, an increased positive correlation of Pow2b (Not) with the PoW-wows of the other group (Pow9, Pow10, Pow18) is observed. Interestingly, Pow2b is now less correlated to Pow9 (Known).
- An increased positive correlation of Pow9 with the PoW-wows of the Not group (Pow1 and Pow2a) is now observed. Moreover, we observe a decrease in its correlation with Pow10 and Pow18.
- Finally, Pow18 is now negatively correlated to both Pow1 and Pow2a.

We repeated the above computations by standardizing the variable values by z-score, as shown in Table 23-4.

Table 23-4. Proximity matrix.

	Pow1: C1	Pow2a: C1	Pow2b: C2	Pow9: C1	Pow10: C2	Pow18: C2
Pow1: C1	1.000	.987	-.999	.869	-.921	-.993
Pow2a: C1	.987	1.000	-.977	.778	-.845	-.999
Pow2b: C2	-.999	-.977	1.000	-.894	.939	.986
Pow9: C1	.869	.778	-.894	1.000	-.993	-.808
Pow10: C2	-.921	-.845	.939	-.993	1.000	.870
Pow18: C2	-.993	-.999	.986	-.808	.870	1.000

According to the statistical computations indicated in Table 23-4, the PoW-wows fall into the following two clusters:

- Cluster 1: Pow1, Pow2a, Pow9
- Cluster 2: Pow2b, Pow10, Pow18

By re-computing the box-plot representation of this new clustering we can observe the significant separation between variables under study for the two groups (see Figure 23-2).

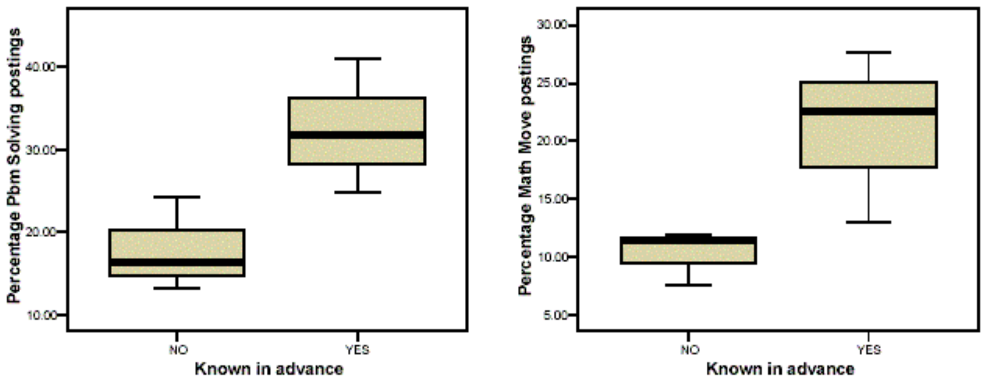

Figure 23-2: Box-plots of problem-solving and math-move dimensions.

In other words, we expected the chat logs to be clustered based on the idea that in some chats, participants had access to the problem prior to their participation in the chat, while in other chats, participants had no access to the problem. However, the statistical analysis demonstrated that the clustering of chats was organized according to some other basis. At this point, we determined to conduct a qualitative approach to identify the reasons for this alternative organization of the online chats.

The Conversation Analysis

To discover possible reasons for the failure of our initial hypothesis that the six PoW-wows would fall into two clusters based on the Known/Not criterion, we re-examined the chats using CA. We decided to see if we could find a difference in participation frameworks organized by the students in the two clusters. For this approach, we examined logs of the online chats to identify participants' perspectives on their own actions, with an eye to describing their actions as sense-making procedures by which they distinctively organized their interactions and their participations in the chats.

The work of conversation analysis involves close inspection of interactional data. In conventional face-to-face interaction, this involves inspecting video and audio recordings of interaction (including non-verbal glances, gestures, facial expressions and bodily orientations, as well as verbal hesitations, repeats, silences, intonation, etc.). When it comes to online chats, the data inspected are just the textual logs of the chats, which display the text postings of participants, the participant's handle (login name) and the time stamp associated with each posting.

The object of inquiry in conversation analysis is not exclusively conversation per se, but rather talk and social interaction. Thus, as Ten Have describes it, "CA's interest is with the local production of [social] order and with 'members' methods' for doing so" (1999, p.19). As Psathas writes,

> Conversation analysis studies the order/organization/orderliness of social action, particularly those social actions that are located in everyday interaction, in discursive practices, in the sayings/tellings/doings of members of society. (Psathas, 1995, p.2)

Using the methods of CA, we began to notice that the organization of social order in these chats could be differentiated according to the way that participants oriented to the production of problem solutions. In particular, we noticed that, in some circumstances, participants reported on work they had already completed, whether it was work done prior to the chat or work done offline and without the participation of others in the production of that work during a chat. This organization of participation we have termed "expository" participation. On the other hand, we noticed that there were circumstances in which participants engaged each other (as a group) both in the investigation of the problem and in the production of possible solutions. This organization of participation we termed "exploratory" participation.

Expository participation in the chats we examined involved one actor producing a report as an extended narrative of an activity performed by that actor. Such reporting is designed to project recipient participation in terms of the production of assessments of the report or the reported work. Recipients of that report have not participated in the work being reported. The report is designed and presented either as an already achieved understanding of the problem in terms of a candidate solution or as steps anyone with appropriate understanding of the problem might take to produce a solution.

One version of expository participation is where one actor first announces that a solution has been achieved and then, upon prompting from recipients, proceeds to tell recipients what the solution is and how he or she produced the candidate solution. For example, in the chat excerpt from Pow2b reproduced in Log 23-1, the student named AH3 reports: "**I think I have the solution!**" This calls upon the recipients of this message (the other students in the math team) to solicit the result. REA asks "**what**" the solution is that AH3 found. To this solicitation, AH3 offers, "**The solution is sqrt(74).**" Announcing a result makes it relevant for recipients to ask for an explanation. REA then asks "**how**" AH3 arrived at that solution. Explanations might be offered in ways that describe the production of the solution as having been already achieved by the actor reporting the result, as in, "**First I did ... and then I computed ... which equals ...**" Another way to produce an explanation involves the circumstance where an actor describes how a competent person would go about solving the problem, as in "**First you do ... then you compute ... which equals ...**" In this regard, these approaches to the exposition of a problem's solution is much like the telling of a story (see, e.g., Sacks, 1962/1995). AH3's exposition, consisting of a series of seven uninterrupted postings (34-40) points his teammates to a formula given on a Math Forum site as a resource needed for understanding his solution to the problem. In school math, the whole trick of solving a problem is often selecting the standard formula to apply. The expository character of the chat consists of exchanges like announcement/solicitation, solicitation/report, report/question, question/explanation, which drive the group interaction, along with extended turns being granted to the expositor without undue interruption from the rest of the group.

Log 23-1.

24	**AH3**	**I think I have the solution!**
25	**REA**	**what**
26	**MCP**	**I guess 15**
27	**REA**	**k**
28	**MCP**	**I think it's like the Pythagorean idea, applying to triangles.**
29	**AH3**	**sqrt(5^2 + 7^2) = sqrt(74)**
30	**MCP**	**Yes, 30-60-90 is needed fact**
31	**AH3**	**The solution is sqrt(74)**
32	**REA**	**how**
33	**MCP**	**7?**
34	**AH3**	**Go to...**
35	**AH3**	**http://mathforum.org/dr.math/faq/form ulas/faq.triangle.html**
36	**AH3**	**Under scalene triangle, the formula for the area of any triangle is...**
37	**AH3**	**K = a^2 * sin(B) * sin(C)/[2 sin (A)]**
38	**AH3**	**Why is that smiley their**
39	**AH3**	**K = a^2 * sin(B)* sin(C)/[2 sin (A)]**
40	**AH3**	**Where a = an edgelength of an isosceles triangle**

An expository report is a way that an actor constitutes a problem as solvable. This characterization can be supported because there is evidence in the transcripts that actors themselves orient to these reports in just this way. For example, the actor producing the report treats the problem as having already been solved and thereby constitutes a participation framework in which he or she acts in the manner of an instructor, explaining what is already known by the instructor to an audience that presumably does not yet know. Constituting such a participation framework is a delicate business in the conduct of these chats—partially because within a peer group it positions the explainer as an authority and the others as lacking knowledge. To do so, actors often draw upon the resources of news reporting by indicating they have something newsworthy to report, i.e., the solution to the problem. In other words, it is not that they possess knowledge that makes them superior, but that they have discovered something and are just pointing it out to others. The actor reporting the solution designs his or her report in a way that allows the recipients of the report to "discover" for themselves in the report how the problem can be seen as solvable and solved. Thus, e.g., AH3 just points the others to a resource that is available to all and allows them to work out the solution themselves to compare with the result that AH3 had discovered in the same way.

Exploratory participation, on the other hand, is a more explicitly egalitarian peer process. It involves group participation patterns in which actors interact so as to constitute, in and as their chat, an understanding of a problem in terms of the conjoint (group) production of possible organizations of mathematical activity from which a solution could be achieved. In such circumstances, actors use the resources afforded them by their interaction to constitute the math problem and their understanding of that problem as an emergent sequence of possible and/or achieved math activities designed to produce what may come to be subsequently recognizable and treated as a solution to the problem. If expository participation is a form of "news" reporting, then the distinguishing feature of exploratory participation is that the actors themselves are constituting the "news" as their ongoing interaction rather than reporting it and receiving the report.

Actors engage in exploration by identifying and offering candidate formulations of the problem and possible solutions by constituting and drawing on resources, which are distributed among participants and which are made available by actors' participation in the chat. Like expository participation, the work of exploratory participation also constitutes the problem in terms of its solution, but with exploratory participation neither the solution nor the problem itself are treated as settled matters by participants. Exploratory interactions involve putting forward proposals for consideration and assessment, negotiating ways of formulating the problem in terms of different solution strategies, soliciting resources, candidate solutions, versions of the problem and so on from other participants. Thus the work of exploration often involves articulating alternative provisional versions of the problem in terms of the development, presentation and assessment of possible knowables, as well as alternative possible solutions for the purpose of identifying a problem participants can work on. This is shown in Log 23-2, also excerpted from Powwow1.

Log 23-2.

47	GOR	what's the question
48	PIN	how long is CE
49	PIN	well isnt AB proportional to DE
50	REA	maybe
51	GOR	what's the question
52	REA	are the two similar
53	REA	and how
54	PIN	maybe by Angle Angle
55	PIN	angle C by reflexive
56	MUR	please refer to http://mathforum.org/pow/vmt/feb1204/problem.html for the question GOR.
57	PIN	and then angle A is congruent to angle D
58	REA	hold up
59	PIN	becuase corresponding angles congruent?
60	PIN	if lines are parallel, corresponding angles are congruent
61	REA	true
62	GOR	what's BC
63	PIN	doesnt say
64	PIN	well look, we know the 2 triangles are similar
65	PIN	lets see if we can do anything from that
66	REA	lte's say y is DC
67	REA	and x is CE
68	REA	so if similar ; 5/8= y/AC= x/BC
69	REA	it is proportional
70	PIN	yep
71	GOR	there's two variables

72	REA	yes there is
73	GOR	so what are we going to do now
74	REA	That means 2 or more equations
75	PIN	ya
76		<GOR has left the room.>
77	REA	PIN
78	PIN	ya
79	REA	I have the idea of solving the question
80	PIN	what ya thinkin?
81	REA	If we some how get angle b congruent to angle c. Then triangle DCE is isosceles
82	REA	so if DE is 5
83	REA	then CE has to be five
84	REA	This could include sin, cos, and tan
85	PIN	im thinking that we have to use the other info they gave us
86	PIN	abotu the bisector
87	PIN	some how
88	PIN	cuz then why else would they put it?

In this segment, actors are producing what they themselves take to be incremental displays of both the problem and candidate solution steps as proposals to be taken up and assessed by others for how they might contribute to the production of a solution. In such circumstances, participants' postings constitute their epistemic stance with respect to the material presented. Epistemic stance displays the participants' orientation to the 'truth' value of the propositions being put forward. Actors often use explicit markers like "**I think that ...**," or "**It could be that** ...," etc., as a way of producing mitigated or less that fully committed positions concerning their degree of certainty with respect to propositions they put forward in interaction. However, explicit markers are not always required, especially if the participation of recipients of a proposition is organized in ways that make them responsible, at least in part, for determining the appropriate epistemic stance to take with respect to the posting.

In this example, GOR is a latecomer to the interaction and is soliciting a version of the problem from participants (lines 47 and 51). The moderator, MUR, refers this participant to an online location where the problem statement can be found. In the meantime, PIN and REA engage in an exploration of possible approaches to the solution to the problem (lines 48-50; lines 52-61; lines 64-75). At line 79 REA announces the she might have a solution strategy. This was put forward as a possibility for consideration and assessment, not as an account of previously achieved accomplishment. This description is produced specifically to display to PIN that (a) the status of REA's candidate solution is less than certain (Kärkkäinen, 2003; Pomerantz, 1984), and (b) that PIN is called on to assess the epistemic status of the proposal. PIN's response at lines 85-88 suggests that some of the information provided in the problem statement constitutes a resource that must be considered and incorporated into any solution approach they might derive. The mere fact of the presence of the information in the problem description in the first place provides for its relevance as part of the solution. This contrasts with an expository approach where the speaker would not propose but rather would *report on* a solution step in a way that did not require the participation of others in the interaction to affirm the certainty with which it is presented. Actors' participation and the way they constitute their propositions in the ongoing interaction are thus fundamentally different between expository and exploratory organizations of interaction.

It is important to note that expository and exploratory work may be done during the same chat. Furthermore, expository participation requires that the expositor did the work of producing a solution "offline," i.e., without the participation of other actors in the chat. One of the affordances of chat is that such "offline" activities are possible even as a chat is occurring. Participants only have access to the messages that are posted. An actor's work with a pencil and a pad of paper beside his or her computer is not available to others unless and until it is posted in the chat system for others to inspect and assess.

Solving the Puzzle

By examining the PoW-wow chats, we were able to see that there were qualitatively significant differences in the way participation was organized. Despite the fact that actors in Pow2b had not seen the problem in advance of their chat, they did their work "offline" during the chat and displayed an expository organization of participation—in common with Pow10 and Pow18. Despite the fact that the actors in Pow9 had access to the problem in advance of the chat, they displayed an exploratory organization of participation—in common with Pow1 and Pow2a. Thus, using CA, we were able to identify the same correlation among the PoW-wows discovered by the statistical analysis. Moreover, whereas the clustering of the PoW-wows flew in the face of the statistical hypothesis, the conversation analysis provided a clear explanation of the clustering in terms of the organization of participation in the chats.

Once we solved the puzzle that emerged from the statistical analysis, we considered whether or not our coding scheme could have been used to identify these different organizations of participation—through a different analysis. We decided that it would not have been possible. The primary reason for this decision was that the existing coding scheme treated the individual posts as the primary units of analysis. Codes applied to individual chat postings could not be used to characterize larger sequences of postings. This made it impossible to analytically identify the organization of participation, understood as a relation among groupings of posted chat messages.

While an alternative coding scheme defined at a different unit of analysis might have made an analysis of exploratory vs. expository organization possible, it would have raised a logical problem of consistency: that the use of coding schemes is generally conducted in ways that lend themselves to finding things for which there are codes. I.e., to distinguish exploratory from expository chats, we would have to design codes for characteristic features of these chat forms. We concluded that if we want to understand how participants organize their participation—if we want to understand a sequence of actions from the participants' perspectives—then the coding scheme would need to capture these perspectives rather than a preconceived (*a priori*) perspective or interest of the researcher.

While we found coding problematic from a CA perspective, we recognized the need for quantitative measures for certain kinds of important claims that we would like to be able to make. According to Heritage & Roth (1995) practitioners of CA have often made informal distributional claims with respect to observed interactional phenomena—e.g., that certain methods of accomplishing interactional tasks are typical, at least within specific linguistic communities. However, questions about the typicality or distribution of certain features of interactions of a particular type can ultimately only be measured quantitatively. We need a way to classify (code) interactions (at some appropriate unit of analysis) so that they can be counted and compared to similar counts from contrasting sets of interactions. In such cases, questions arise as to the appropriate way to code data such that the requirements of valid statistical and quantitative analysis can be met without violating the requirements of preserving the participants' perspectives on the sequential organization that they create in their unfolding action. In order to determine whether our qualitative results provide an adequate explanation across multiple cases, we need to re-specify a coding scheme that derives from the perspective of the participants, as observed in our logs (for further discussion, see Heritage & Roth, 1995; Kaplan, 1964).

As explained in the remainder of this chapter, we have begun to explore an approach to coding, based on the ways that interactants organize themselves and their interaction into recognizable activities. This approach uses CA methods to identify closings and openings of action sequences, by which participants organize their activities into "long sequences" (Sacks, 1962/1995) of identifiable action types. For example, we have begun to identify sequences in which math problem-solving activities are being conducted, as distinct from various other kinds of non-math social interaction. In this way, we are developing a coding scheme that preserves actors' orientations, concerns, relevancies and their sequential organization of the ongoing interaction. This proposed approach to coding makes possible the comparison of different instances of social interaction in ways that preserve the participants' organization of interaction and exploit that local organization as a source of insight into the ways we come to treat action sequences as sequences of particular sorts.

A CA Approach to Coding

In conducting inquiry into matters of collaboration, learning and instruction, the analyst is confronted with a range of methodological and assumptive commitments that shape the nature of the research performed and the kinds of claims that can be sustained by that research. In examining the early VMT chats, we considered how best to begin asking relevant questions from a CA perspective. For example, we were concerned with questions such as, "What are the chat participants *doing* in these chats?" "Are their chats collaborative and, if so, *what* makes their chats collaborative?" "How do these students *organize* their interactions?" "How do these students *do* math in an online environment?" "Are there similarities and differences in the way these chats are *done*?" The rest of this chapter represents our effort to conduct an analysis based on assumptions from the CA perspective about human action, social interaction, collaboration and communication.

In this research, we have begun to develop a CA-informed alternative to classical SA coding. Our approach is based on ethnomethodological assumptions regarding sense making, action and the competence of participants. The main difference between the two approaches consists in the definition of the "unit of analysis": while in the SA approach the unit of analysis is chosen by the analyst (usually a unit of fixed length or a posting), in the CA approach it is identified according to the participants' perspective within the interactional situation.

Since postings are authored and contributed by each participant, one can argue that selecting individual postings as the unit of analysis is not an arbitrary choice imposed by the researcher on the data. Indeed, at first glance this seems to be a natural choice, which is compatible with the participants' perspective. However, the arbitrariness of this choice becomes clear when one thinks about the interactional work that each posting is designed to accomplish in chat. The quasi-synchronous nature of the environment and the fact that one needs to type his/her contributions encourages participants to interact with each other in particular ways in a chat environment. Participants are pressed to quickly submit multiple short texts in order to post their contributions at relevant points. Due to this characteristic of the chat interface, it is often the case that only a combination of postings constitutes a coherent turn or activity. More importantly, an individual post is essentially situated within the larger context, in which it must be understood as a response to previous activities or at least to the possibilities opened up by them. It must also be seen as a solicitation of responses and follow-ups, or at least as a text designed to be understood by other participants, who may be expected to then express their comprehension or lack thereof. Thus, as far as an analytical effort that aims to study the organization of collaborative activities in a chat environment is concerned, considering a single posting as the unit of analysis without making any claim about its relationship to other activities would be a premature choice.

We used CA methods to identify how the chat participants themselves organized their interaction into "long sequences" (Sacks, 1962/1995) or, as we call them, "chunks" of activity. The VMT research team engaged in numerous data sessions to identify those locations in the chat where new activities were initiated and where ongoing activities were suspended or brought to some kind of closure. In so doing, we were able to identify activity sequences to which the participants themselves visibly oriented. An activity sequence in this sense is a set of postings that are highly connected in terms of their response structure and that work together to accomplish some coherent activity that can be observed (by the participants and the analysts) in the design of the postings as the focus of the postings. For instance, the explicit and implicit indexical references of the postings tie the individual postings together as contributions to the activity. We then assigned labels to these activity sequences based on the way the participants themselves oriented to, conducted and regulated their actions in these activity sequences. In so doing, we were able to identify how participants themselves managed the sequential organization of their math problem-solving chats. We were also able to apply our labeling schemes across the six different chats discussed in the beginning of this chapter (see Table 23-1), making it possible for us to begin to compare how these chats were organized.

We base our approach on the presumption that the sequential organization of the interaction is the basis by which participants and observers alike make sense of the online collaborative activity of a small group of participants. We call this approach "participant-centered analysis." The basic idea is that the perspective of the participants and the work they do to make sense of their own actions provides the ground for organizing their interactional work into coherent long sequences or chunks of activity. By incorporating participants' perspectives and trying to get a sense of the organization of their activity in terms of the ways they themselves achieve that organization, we hope to demonstrate that it is possible to begin to do quantitative analysis in ways that do not elevate the analyst or privilege the analyst's perspective over the perspectives of the participants. We firmly hold that the sense and coherence of interaction is locally produced for and by the participants in that interaction. This suggests that the analyst's role is not to impose an external sense-making structure based on some theoretical interests of the analyst on observed activity but to allow the participants' own sense-making work to become evident and to allow that sense-making work to reveal itself in the coherent ways that ongoing action is organized and produced.

Though our approach differs substantially from the classical SA approaches, there are some similarities such as the use of labeling/codes and a sort of multilevel approach. It is worth mentioning here that ours is a top-down approach, starting from "high"-level activities in which the participants engaged, to the most detailed levels of interaction shown/found in the data—beyond the level of the posting to individual lexical, syntactical and indexical features. In contrast, in classical coding the multilevel approach is done in a bottom-up fashion (from a single posting to groups of postings representing "activities"), in order to look at the distributions of the codes, and consider aggregations and vector representations derived from these values to do hypothesis testing and comparisons. Because our CA approach takes this top-down view, we do not *reconstruct* the activities—as it is done in the classical SA approach—but use the organization of activities achieved by the participants themselves as a way of conducting analysis. Although there are some SA studies that focus on the sequential relationships between codes to make claims about the type of the ongoing activity at certain episodes (e.g., whether a given episode is an effective knowledge-sharing episode), they usually assume a simple linear ordering of short sequences of postings (e.g., Soller & Lesgold, 2003).

Long Sequences

While most CA research examines very short sequences of interaction (such as adjacency pairs and their elaborations), long sequences have also been matters of concern for conversation analysts. Sacks (1962/1995) devoted a lecture to long sequences and remarked:

> A basic sort of investigation is that of long sequences as a coherent matter as compared to simply studying utterance by utterance, a long sequence which you then have as an in-some-way connected series of small fragments. And such investigation is, if it's going to develop at all, at a rather primitive stage—leaving aside obvious sorts of things where you're dealing with relatively game-like situations or other sorts of known, pre-organized matters. The sequences we're dealing with are not pre-organized. (Vol. II, p. 355)

As Sacks noted, conversation analysts have developed an extensive body of research regarding the observable regularities, those “series of small fragments” that are produced in the conversations they constitute. But the issue of long sequences, packages or chunks is of a different order. Sacks recognized that chunks were not simply assemblies of smaller sequences:

> Certain aspects of the work you might do on a small sequence won't do you any good in trying to package longer sequences. Indeed, they might be misguiding in that you would figure that you've dealt with some pair in some fashion, and even in a sequential fashion, and thereby not see the potentiality for building a larger package for which the way you had studied the smaller sequence didn't have much bearing, or had only some relatively intricate bearing. (Vol. II, p. 354)

In fact, the classical object of conversation analytic interest, i.e., the conversation, is actually a gloss for a kind of long sequence of social interaction involving something like informal talk, i.e., multi-turn, multi-participant interactions that are not pre-organized, that are composed of sequences of talk, gesture and other forms of embodied action, and that are built to be and treated as coherent by the participants who produce them (Schegloff, 1990). Recent work has also begun to investigate features of other kinds of long sequences like the medical interview (Maynard, 2003), negotiations (Firth, 1995), talk at work (Boden, 1995; Suchman, 1987) and different organizations of institutional discourse (Drew & Heritage, 1993). These studies all treat long sequences as locally situated and contingent achievements that are organized and produced in ways that allow participants to treat their participation in them as participation in ongoing and contingently coherent activity.

Among the regularities observed and studied by conversation analysts are the ways that long sequences begin and end. For example, participants in conversations engage in recognizable boundary-producing activities to which participants orient and by which participants initiate conversations and bring them to a close. These are referred to in the literature as *openings* and *closings* (Schegloff & Sacks, 1973; Schegloff, 1968). These kinds of activities are also used within conversations as ways that participants display to each other that some activity in which they had been engaged is completed or suspended and another is starting. As such, they serve to mark something like boundaries between long sequences in an ongoing interaction and allow participants a wide range of opportunities to manage, regulate and build their interaction to become coherent stretches of lengthy activities.

Upon close examination of the VMT PoW-wow chat transcripts, it became apparent that the participants themselves were orienting to and organizing their participation in the chats in terms of long sequences of interaction that extended beyond conventional conversation analytic notions of the turn and the adjacency pair

(Schegloff, 1990). Participants organized their interaction into longer sequences, sequences that were coherent by virtue of their sequential organization, by virtue of the fact that "participants are oriented to finding coherence—'if they can'" (p. 73). According to Schegloff, the coherence of these long sequences is a structural feature of the way they are opened, expanded and closed.

As a practical analytical matter, we began with the following noticings in order to identify these longer sequences in the chats. First, the chats were of finite duration; they had identifiable beginnings and endings, which the participants themselves performed and to which they oriented as relevant in the conduct of their chat. In addition, participants appeared to organize their interaction into long sequences in which they attended to the math problem, worked out problems associated with the distribution of geometric figures to other chat participants, dealt with problems associated with the chat technology itself, and engaged each other with respect to matters other than the math problem they were discussing.

These noticings led us to consider how we might be able to distinguish among these long sequences or chunks of activity across a number of different chats to see what, if any, similarities or differences there might be in the way that chats were organized by the participants themselves. To achieve this, we first elected to use CA methods to identify actions such as openings and closings of various sorts that indicated participants were initiating, suspending and/or closing a sequence or chunk of activity. This allowed us to identify coherent long sequences of activity based on participants' own methods of organizing their activities. We then created a participant-centered coding scheme by assigning labels to the chunks of activity we identified, producing what effectively might be termed a data dictionary. Finally, we were able to develop visual representations of these chunks of activity and draw certain conclusions based on this participant-centered coding scheme.

Identifying Long Sequence Boundaries

The approach we are developing in this research is a form of participant-centered analysis (PCA). PCA involves identifying and working with those features of social interaction that the participants find relevant and to which the participants orient in their ongoing participation in the social interaction. To do this, we inspected six transcripts of the VMT PoW-wow chats in detail using CA methods to identify how the participants in the activity had organized their activity. One way we did this is to identify openings and closings by which participants either (1) bring one activity to a close and initiate another activity or (2) suspend an ongoing activity and initiate a new activity. In Log 23-3 from Pow2M (referred to as Pow2b earlier), we see two such transitions. One begins at lines 10 and 11, and the other begins at lines 21-23.

Log 23-3.

5 MUR Hi. Thanks for participating in our PoW-Wow. For privacy reasons, we're asking that you don't share any personal information about yourself, such as your name, age, or where you live.

6	MUR	Let's go around and have everyone share a greeting with the group. I'll start by saying that I'm really looking forward to seeing you talk about math tonight!
7	AH3	Hello everyone
8	REA	Hi
9	MCP	Hi! Last time was fun, and I look forward to this --
10	REA	I remember you MCP
11	MUR	OK, here are four guidelines that we'll use tonight.
12	MUR	1. During the session, share ideas about how to solve the problem.
13	MUR	2. Feel free to ask about anything that seems unclear.
14	MUR	3. If you all think the problem is solved please make sure everyone in the group understands the explanation for the answer.
15	MUR	4. I'm here if you have any technical problems or questions, but I won't help with the math.
16	MUR	Here's the problem that you'll be working on tonight:
17	REA	where is the problem
18	MUR	If two equilateral triangles have edgelengths of 9 cubits and 12 cubits, what's the edgelength of the equilateral triangle whose area is equal to the sum of the areas of the other two?
19	MUR	You can also read the problem at http://mathforum.org/pow/vmt/feb2604/problem.html
20	MUR	Good luck:-)
21	MUR	By the way, if you create a picture that you would like to share with your group, there are instructions on the problem page about how to do that.
22	MCP	Probably a straight area compute. B4 I do it, I want to guess, ok?
23	REA	Have any of you guys learend of the 30-60-90 concept
24	AH3	I think I have the solution!
25	REA	what
26	MCP	I guess 15
27	REA	k

Upon examining this fragment of the transcript, it became evident that the posting at line 11 was designed to do two things: initiate a new activity and close down the prior activity. In particular, the use of the particle "OK" in line 11 is specifically designed to indicate both an opening and a closing (Beach, 1993; Condon, 2001). In this usage, "OK" is a transition marker designed to indicate that a new (and as yet unspecified) activity is about to be initiated. In so doing, it also serves to bring to a close the prior interaction.

Another transitional moment occurred at lines 21-23. In particular, line 22 displays uptake by a student of the problem identified in the previous set of instructions and marks the close of the sequence of instructions for doing the problem (lines 11-21), which are provided by the facilitator, MUR. This uptake is affirmed by REA's subsequent post (line 23), in which he addresses a problem-relevant question to other participants. Thus, for our purposes, we did not consider the "content" of MUR's posted messages in lines 11-21 to identify it as a long sequence or "chunk of activity." What makes this segment a chunk is the fact that it has a discernable opening produced and taken up by the participants and a discernable closing, which is also produced and taken up by the participants.

One of the features of chats is that strict adherence to conversational turn taking is problematic for participants (see Chapter 21). Thus, it is often the case that a person produces a post that is in response to some prior post other than the immediately prior post. This is an artifact of chat technology, which makes it possible for two different activity sequences to be "interleaved," as Log 23-4 from Pow1 demonstrates.

Log 23-4.

81	REA	If we some how get angle b congruent to angle c. Then triangle DCE is isoceles
82	REA	so if DE is 5
83	REA	then CE has to be five
84	REA	This could include sin, cos, and tan
85	PIN	im thinking that we have to use the other info they gave us
86	PIN	abotu the bisector
87	PIN	some how
88	PIN	cuz then why else would they put it?
89	REA	two ideas
90	MUR	We have a new participant who wants to join you. Do you mind?
91	PIN	fine w/ me
92	REA	tnope
93	REA	nope
94		<MCP has entered the room>
95	REA	never mind about the two ideas
96	PIN	k
97	MUR	Hi MCP. Could you guys help MCP to catch up?
98	PIN	sure
99	REA	k
100	MCP	I just read the prob and got a diagram.

Prior to line 90, student participants had been working on finding a solution to the math problem on which they were working in this chat. In lines 85-88, PIN had problematized "**the other info**" made available in the problem in a particular way. PIN's remarks were designed to make questionable to and for other student participants what could serve as an adequate account for the availability of that "**other**" information in the first place. REA responded to PIN's solicitation of an account with a prefatory posting at line 89 that indicated that a subsequent expansion of what those "**two ideas**" were would be forthcoming as a next set of postings from REA.

One of the features of this chat is that work on the math problem was done by students. The facilitator served only to regulate certain aspects of their interaction (i.e., introduce newcomers to the chat) and attend to technical questions (i.e., methods for disseminating drawings of the problem to chat participants). So, when the facilitator announced that there was another participant who wanted to join the chat at line 90, the very appearance of a post from the facilitator indicated that something other than the problem-solving work the students had been engaged in was about to begin. MUR's posting at line 90 calls on participants to suspend their ongoing work on the problem and to indicate their willingness to accept a newcomer to the chat. PIN and REA indicated their willingness in lines 91 through 93.

Line 94 is a system-generated message indicating that the new participant, MCP, had entered the chat room. At this point, REA posts a message, the sense of which is derived from the problem-solving work they were doing immediately prior to MUR's intervention at line 90. Specifically, REA proposes to close the prior discussion of reasons for the additional (and as yet apparently unused) information provided in the problem statement at line 95. PIN accepts this proposal at line 96, bringing to a close (at least for the moment) any further consideration of the problem. This is followed almost immediately (after one second) by a greeting from MUR and by MUR's request that the other students bring MCP current with their work on the problem. Thus, we see that interleaved in MUR's opening intervention is work done by student participants that is relevant to closing the problem-solving work in which they had been engaged prior to the intervention.

The Data Dictionary and Long Sequences

The preceding instances serve as examples of the way that CA methods were applied to identify boundaries between long sequences. The next step was to apply these methods to the six chat transcripts and to identify those postings belonging to each of the long sequences that formed the chat. We derived a set of descriptive labels for the long sequences, which served as a provisional data dictionary for this first level of long sequence analysis. These are shown in Table 23-5.

Table 23-5. Data dictionary.

Code	Explanation
STARTCHAT	STARTING THE CHAT
FACLn	FACILITATOR GIVES INSTRUCTIONS, NUMBER n
PBn	PROBLEM-SOLVING SEQUENCE, NUMBER n
PICn	SEQUENCE INVOLVING POSTING OF PICTURES, NUMBER n
CATCHn	PARTICIPANTS WORK TO ALLOW ANOTHER TO CATCH UP WITH THE WORK THAT HAS BEEN DONE, NUMBER n
SOCn	SOCIALIZING SEQUENCE, NUMBER n
SERVICEn	SEQUENCE CONCERNING MATHFORUM SERVICES, NUMBER n
ENDCHATn	SEQUENCE TO END CHAT, NUMBER n
LOSTn	STUDENTS DEALING WITH PARTICIPANT WHO IS LOST, NUMBER n
ASKHELPn	STUDENTS REQUEST MATH HELP FROM FACILITATOR, NUMBER n
NEWMEM	NEW MEMBER JOINS THE CHAT
QUIT	STUDENT QUITS FROM CHAT
TECH_PBn	ADDRESSING A TECHNICAL PROBLEM WITH THE CHAT, NUMBER n
PAUSE	PARTICIPANT TEMPORARILY SUSPENDS PARTICIPATION IN THE CHAT
SYSMESSAGE	MESSAGE PRODUCED BY CHAT SYSTEM
SYSBREAK	TECHNICAL BREAKDOWN OF CHAT SYSTEM
CH_GP_STATn	CHECKING THAT ALL PARTICIPANTS UNDERSTAND WHAT IS GOING ON, NUMBER n
CH_W_FACn	SEEKING ASSESSMENT OF RESULTS FROM FACILITATOR, NUMBER n
PB_W_FACn	DOING PROBLEM-SOLVING WITH THE FACILITATOR, NUMBER n
RES	PARTICIPANT PRODUCES ACCOUNT OF ACTION

The labels we applied to the sequences were designed as provisional and defeasible shorthand descriptions of the activity performed in the sequence. Other descriptors are certainly possible, but these seemed to be adequate for our purpose of characterizing sequences in terms of what the participants were doing in them. Each long sequence is often composed of smaller sequences, which may be quite long in their own right. For example, doing problem solving involves a number of activities, all of which were grouped together to form our problem-solving sequences.

Graphical and Statistical Analyses

We distinguished our sequences according to the way that participants themselves brought them to a close or temporarily suspended their participation in them and initiated activities that were not related to the work done in that sequence. This allowed us to produce graphical representations of the chats, which showed the sequential organization of the chats in terms of the long sequences of which they were composed. These are shown in Figures 23-3 through 23-8. A number of interesting results emerged from the various descriptive statistics available for these chats. For example, it is evident from these graphs (rows labeled PB1, PB2, PB3 or PB4) that participants in collaborative problem-solving chats spend a considerable amount of time actually engaged in problem-solving activities.

Individual participants are listed with color codes in the figures (as in Pow1, Figure 23-3). If the chats were lengthy (as in Pow2G, Figure 23-4), a scale factor was used to condense the display, which merged individual contributions, making it impossible to represent with colors the participation in the sequences we identified.

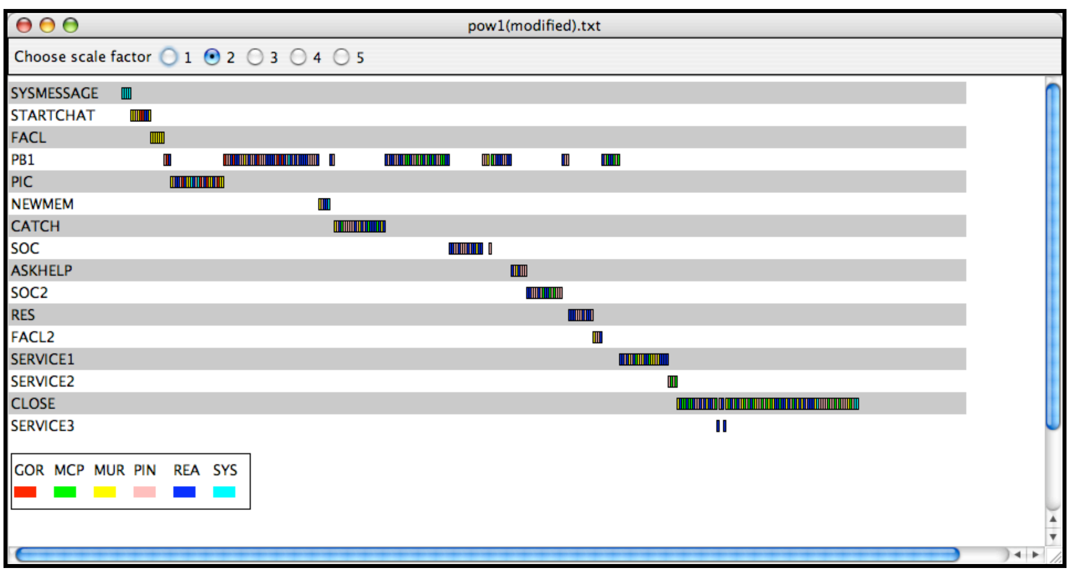

Figure 23-3. Pow1.

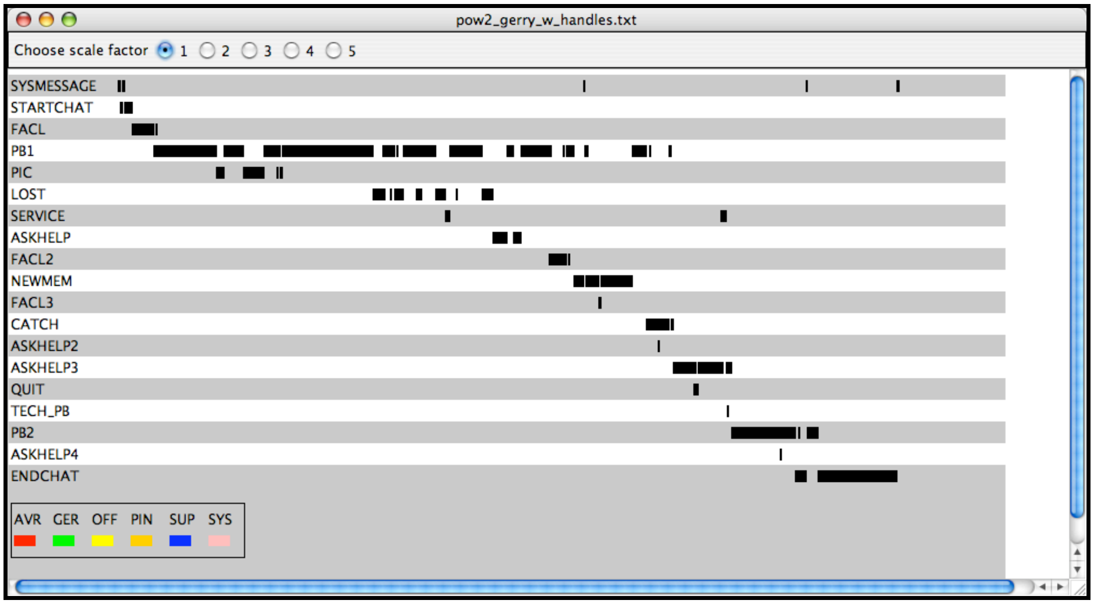

Figure 23-4. Pow2G (referred to as Pow2A earlier).

Figure 23-5. Pow2M (referred to as Pow2b earlier).

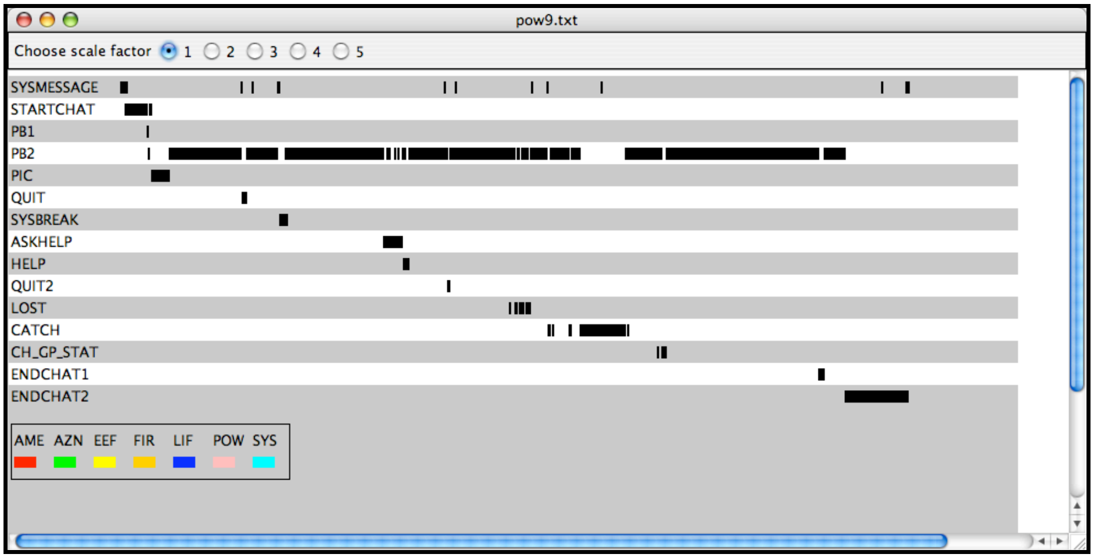

Figure 23-6. Pow9.

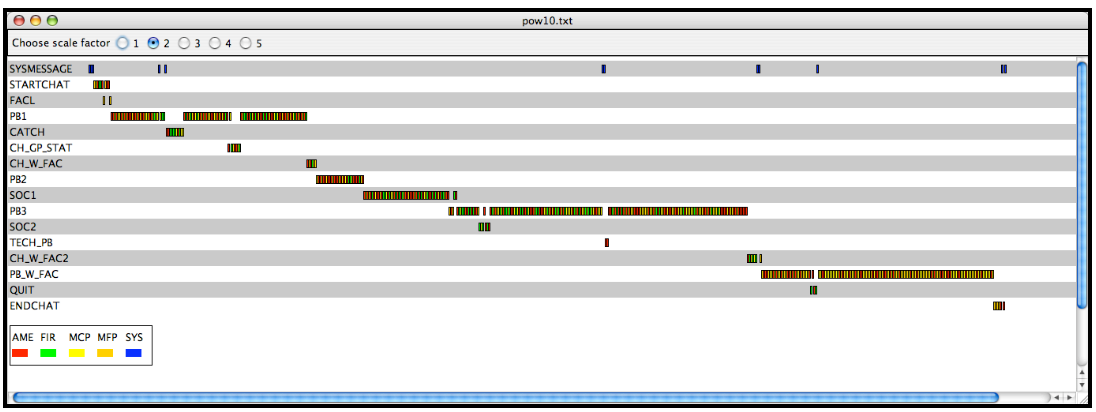

Figure 23-7. Pow10.

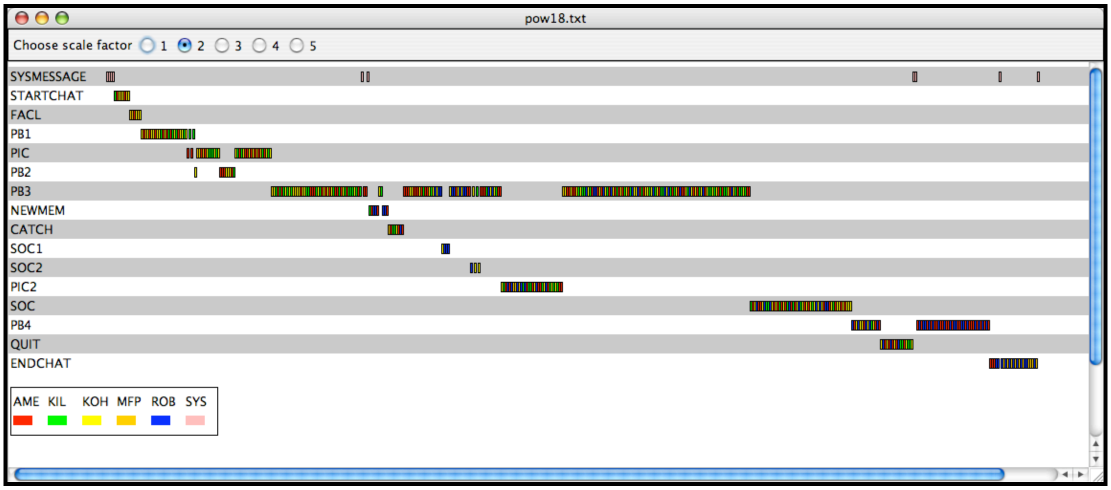

Figure 23-8. Pow18.

The feature of the chats that emerges from an inspection of Figures 23-3 through 23-8—that postings related to problem solving were the most prevalent in each of the chats—is confirmed by the coding statistics shown in Table 23-6:

Table 23-6. Frequency of postings in each activity by PoW-wow.

LONGSEQ * POWWOW Crosstabulation

% within POWWOW

		POWWOW						
		POWWOW 01	POWWOW 02G	POWWOW 02M	POWWOW 09	POWWOW 10	POWWOW 18	Total
LONGSEQ	ASKHELP		9.4%		2.4%			2.8%
	CATCH	8.1%	2.9%		6.4%	1.9%	1.6%	3.7%
	CH_GP_STA				.7%	1.2%		.4%
	CH_W_FAC					2.2%		.4%
	ENDCHAT	23.7%	11.3%	8.9%	8.4%	1.0%	4.9%	8.8%
	FACL	3.0%	5.3%	5.4%		.3%	1.2%	2.2%
	LOST		5.7%		1.3%			1.6%
	NEWMEM		6.8%				1.6%	1.9%
	PAUSE			2.0%				.1%
	PB	29.7%	50.6%	76.8%	71.6%	53.5%	57.7%	56.7%
	PB_W_FAC					24.6%		4.7%
	PIC	6.6%	4.0%		2.2%		13.3%	4.3%
	QUIT		.4%	.5%	.8%	.5%	3.5%	1.0%
	SERVICE	6.6%	1.1%					1.0%
	SOC	16.2%				10.7%	12.3%	5.8%
	STARTCHAT	2.7%	1.2%	3.0%	2.9%	1.5%	1.6%	2.0%
	SYSBREAK				.8%			.2%
	SYSMESSAG	3.3%	1.1%	3.4%	2.4%	2.1%	2.1%	2.1%
	TECH_PB		.1%			.3%		.1%
Total		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

As can be seen in the PB row, on average, problem solving accounted for slightly over half of the postings in the chats (ranging from 30% to 77%). While this may not seem terribly surprising, it is nonetheless quantitative confirmation that the participants themselves oriented to their participation in the chats heavily in terms of problem solving.

Another feature of collaborative problem-solving chats is that student participants were able to organize their chat interaction in ways that allowed them to engage in multiple, concurrently performed activities. It has been claimed that one of the affordances of chat technology is precisely that participants can and do engage in multiple, concurrent activities (Garcia & Jacobs, 1999; O'Neill & Martin, 2003). As the transcripts and Figures 23-3 through 23-8 demonstrate, participants were able to do more than one thing at a time in a number of different ways. For example, actors were capable of suspending their engagement in problem solving over multiple postings to take up a next activity and then return to problem solving where they had left off. Also, they were capable of posting messages in one activity while inserting postings related to a different activity in the stream of current-activity postings.

Among the research questions we asked was the question regarding how similar the different PoW-wows were. We constructed a similarity matrix displaying Pearson correlation coefficients based on the distribution of postings across the categories we had discovered. This is shown in Table 23-7.

Table 23-7. Similarity matrix with all variables.

Proximity Matrix

	Correlation between Vectors of Values					
	1:P01	2:P02G	3:P02M	4:P09	5:P10	6:P18
1:P01	1.000	.736	.731	.741	.630	.779
2:P02G	.736	1.000	.965	.968	.814	.918
3:P02M	.731	.965	1.000	.991	.877	.948
4:P09	.741	.968	.991	1.000	.874	.952
5:P10	.630	.814	.877	.874	1.000	.866
6:P18	.779	.918	.948	.952	.866	1.000

This is a similarity matrix

As the figures indicate, these chats are all quite similar. Again, this is not a surprising result in that the students self-selected to participate in these chats with the understanding that they were going to be doing math problem solving.

We then asked, how similar these chats are with respect to the distribution of postings in non-problem-solving activities. When the problem-solving category was removed, the correlations in Table 23-8 emerged when run with the following variables, ASKHELP, CATCH, CH_GP_ST, CH_W_FAC, ENDCHAT, FACL, LOST, NEWMEM, PAUSE, PIC, QUIT, SERVICE, SOC, STARTCHAT, SYSBREAK, SYSMESSAGE, and TECH_PB.

Table 23-8. Similarity matrix without problem solving.

Proximity Matrix

	Correlation between Vectors of Values					
	1:P01	2:P02G	3:P02M	4:P09	5:P10	6:P18
1:P01	1.000	.376	.559	.638	.493	.582
2:P02G	.376	1.000	.479	.550	-.272	.061
3:P02M	.559	.479	1.000	.547	-.066	.038
4:P09	.638	.550	.547	1.000	-.050	.149
5:P10	.493	-.272	-.066	-.050	1.000	.557
6:P18	.582	.061	.038	.149	.557	1.000

This is a similarity matrix

The idea here was to see how similar chats were with respect to the organization of non-problem-solving activities. As we can see, Pow10 shows small negative correlations to Pow2G, Pow2M and Pow9. The rest show positive but relatively small correlations with other PoW-wows, suggesting that there are similarities in the ways that participants deal with circumstantial contingencies that arise during their chats, but also that there are issues to be investigated with respect to the differences in the kinds of contingencies that arise during these problem-solving chats.

We then did a multidimensional analysis based on the proximity matrices we calculated. Figure 23-9 gives us a graphical representation derived from the similarity matrix in Table 23-9. It appears that the data cluster into three groups:

- Cluster 1, which consists of Pow2M, Pow2G, Pow18 and Pow9,
- Cluster 2, which consists of Pow1, and
- Cluster 3, which consists of Pow10.

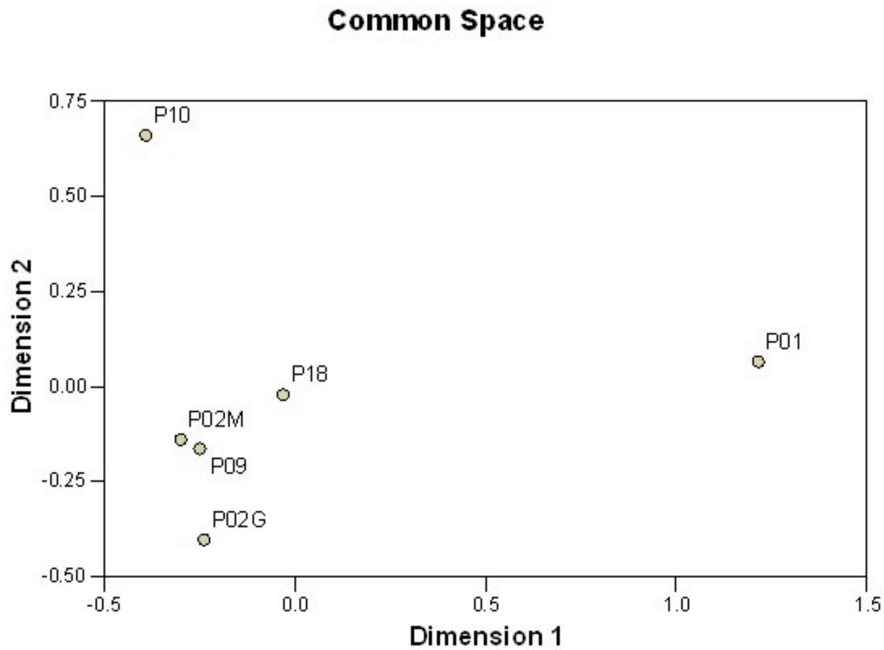

Figure 23-9. Multidimensional scaling analysis of proximity matrix.

Table 23-9. Similarity matrix.

Proximities

	P01	P02G	P02M	P09	P10	P18
P01	.					
P02G	.293	.				
P02M	.280	.927	.			
P09	.308	.936	1.000	.		
P10	.000	.510	.683	.675	.	
P18	.412	.798	.881	.892	.654	.

Goodness of Fit

Stress and Fit Measures

Normalized Raw Stress	.00191
Stress-I	.04367[a]
Stress-II	.07998[a]
S-Stress	.00341[b]
Dispersion Accounted For (D.A.F.)	.99809
Tucker's Coefficient of Congruence	.99905

PROXSCAL minimizes Normalized Raw Stress.

a. Optimal scaling factor = 1.002.

b. Optimal scaling factor = .999.

Some of this cluster pattern may be accounted for by the following. In the first cluster there is usually a main problem-solving activity that is interleaved with other sorts of activities, but the main activity is usually sustained. In Powwow 10 there are not many activities that are interleaved with the problem-solving activity. The activities unfold in a linear way without interleaving with each other. Finally, Powwow 1 lies somewhere in between these two clusters. Except the PB2, CATCH, SOC and PB3 chunks (which add up to almost half of the whole session), the remaining chunks unfold in a linear way without much interleaving. Clearly, further investigation is required to account for the basis for this clustering.

We understand that, from a statistical perspective, we do not have anything like conclusive results. But we do have suggestive results. We see that there are differences in the chats, at least initially in terms of the distribution of activities in which student participants are engaged. Furthermore, we see that there are also interesting structural similarities. For example, Figures 23-3 through 23-8 show that participants are capable of engaging in sustained problem-solving work while dealing with the interactional contingencies that emerge over the course of their chats.

Probability Transition Tables

There are a number of areas we wish to explore further as this research evolves. First of all, we have adopted a top-down approach based on the way that participants themselves organize their activities. This is to be distinguished from the CA work that Heritage & Roth (1995) have done, which uses the ways that participants constitute question-response pairs in presidential news conferences as a basis for doing statistical analysis. Essentially what they did was to look at an activity in which the predominant organization of interaction involved asking questions and offering responses to those questions. We are doing something different. We are looking at a chat among multiple student participants who engage in a variety of different kinds of activity over the course of their chats. We begin by identifying the way the students themselves have organized their interaction in terms of activities to which they were oriented.

As a next step, we want to use CA methods to further characterize the constituent features of these activities. In other words, we are asking questions like, "How are problem-solving activities built?" and "From what kinds of more finely grained activity types are these problem-solving activities built by the participants?" This will allow us to discover the different ways that students *do* problem solving, in terms of how their activity emerges over the course of these chats.

Another area of significant interest is to use the coding scheme we are developing to capture the sequential organization of problem-solving chats. We have begun to develop conditional probability tables (see Tables 23-10 to 23-12) that we hope will allow us to model ways that problem-solving chats are likely to unfold.

Table 23-10. Pow2G probability transition table.

	CATCH	END	FACL	HELP	LOST	NEW	PB	PIC
CATCH	80,95%	0,00%	0,00%	9,52%	0,00%	0,00%	9,52%	0,00%
END	0,00%	97,62%	0,00%	0,00%	0,00%	0,00%	2,38%	0,00%
FACL	0,00%	0,00%	81,58%	0,00%	0,00%	2,63%	15,79%	0,00%
HELP	1,47%	0,00%	0,00%	85,29%	0,00%	0,00%	5,88%	0,00%
LOST	0,00%	0,00%	0,00%	2,44%	73,17%	0,00%	21,95%	0,00%
NEW	0,00%	0,00%	2,00%	0,00%	0,00%	92,00%	6,00%	0,00%
PB	0,82%	0,82%	1,37%	0,55%	3,01%	0,82%	91,53%	1,09%
PIC	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	13,79%	86,21%
QUIT	0,00%	0,00%	0,00%	66,67%	0,00%	0,00%	0,00%	0,00%
SERV	0,00%	0,00%	0,00%	25,00%	0,00%	0,00%	12,50%	0,00%
START	0,00%	0,00%	7,69%	0,00%	0,00%	0,00%	0,00%	0,00%
T_PB	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%
Total:	2,91%	11,77%	5,26%	9,42%	5,68%	6,93%	50,69%	4,02%

Table 23-11. Pow2M probability transition table.

	ENDCHAT	FACL	PAUSE	PB	QUIT	STARTCHAT
ENDCHAT	77,78%	0,00%	0,00%	22,22%	0,00%	0,00%
FACL	0,00%	90,91%	0,00%	9,09%	0,00%	0,00%
PAUSE	0,00%	0,00%	50,00%	50,00%	0,00%	0,00%
PB	3,23%	0,00%	1,29%	94,19%	1,29%	0,00%
QUIT	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%
STARTCHAT	0,00%	10,00%	0,00%	0,00%	0,00%	90,00%
Total:	9,50%	5,50%	2,00%	77,50%	1,00%	4,50%

Table 23-12. Pow18 probability transition table.

	CATCH	ENDCHAT	FACL	NEWMEM	PB	PIC	QUIT	SOC	STARTCHAT
CATCH	87,50%	0,00%	0,00%	0,00%	12,50%	0,00%	0,00%	0,00%	0,00%
ENDCHAT	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
FACL	0,00%	0,00%	83,33%	0,00%	16,67%	0,00%	0,00%	0,00%	0,00%
NEWMEM	10,00%	0,00%	0,00%	70,00%	20,00%	0,00%	0,00%	0,00%	0,00%
PB	0,00%	0,36%	0,00%	1,07%	94,64%	1,79%	0,36%	1,79%	0,00%
PIC	0,00%	0,00%	0,00%	0,00%	7,69%	92,31%	0,00%	0,00%	0,00%
QUIT	0,00%	0,00%	0,00%	0,00%	5,26%	0,00%	94,74%	0,00%	0,00%
SOC	0,00%	0,00%	0,00%	0,00%	8,20%	0,00%	0,00%	91,80%	0,00%
STARTCHAT	0,00%	0,00%	8,33%	0,00%	0,00%	0,00%	0,00%	0,00%	91,67%
Total:	1,65%	5,35%	1,23%	2,06%	57,61%	13,37%	3,91%	12,55%	2,26%

As we do more refined analyses of long activity sequences within the chats, we expect to be able to develop conditional probability tables that describe how such activities as problem-solving or help-seeking activities unfold. While this is not possible at this stage of our research, we feel that with additional work over a larger sample of chats, one would be able to begin to see something about the structural organization of problem-solving chats as interactional phenomena and social facts.

Mixing Methods

One of the most important features of the work we have done here is to demonstrate that very different analytical methodologies can be used together to tackle interesting problems. The key here was to recognize that conversation analysis could be used effectively to provide a coding scheme, based on the interactional relevancies of the participants whose actions were of interest in the study, which could be used effectively to do comparative statistical analysis of the data. Statistical studies often treat anomalies in the data and in findings as random occurrences. Occasionally they are. But rather than assume the status of such anomalies, we took the approach that our initial analyst-based coding scheme might have been responsible for such anomalies. By working to produce a coding scheme based on the demonstrated relevancies of the participants in the interactions under examination, we were able to resolve the anomalies and achieve insights into the data that would have otherwise gone unnoticed.

There are many advocates for mixed-methods studies. We are among them. However, we hold to the position that mixing methods can only be done effectively when analysts give careful consideration to the assumptions governing the organization of all methods deployed, making sure that no method violates the assumptions governing the use of another method. In our examination of the literature, we have found that conversation analysis has not yet managed to develop methods for examining long sequences. On the other hand, statistical analysis is often used to test analysts' hypotheses without regard for the inherent organization of interaction based on participants' practices. However, together, CA and SA can be used to explore the structural and sequential organization of participants' own actions over long sequences and across distinct interactional occurrences in ways that respect the inherent orderliness of the data while allowing for generalization beyond specific instances.

References

Avouris, K., & Margaritis, F. (2002). *A tool to support interaction and collaboration analysis of learning activity.* Paper presented at the CSCL 2002, Boulder, CO.

Beach, W. A. (1993). Transitional regulations for 'casual' "okay" usages. *Journal of Pragmatics, 19*(4), 325-352.

Boden, D. (1995). *The business of talk: Organizations in action*. London, UK: Blackwell.

Condon, S. L. (2001). Discourse ok revisited: Default organization in verbal interaction. *Journal of Pragmatics, 33*(4), 491-513.

Daradoumis, T., Martínez, A., & Xhafa, F. (2004). *An integrated approach for analysing and assessing the performance of virtual learning groups.* Paper presented at the CRIWG 2004. Proceedings pp. 289-304: Springer.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). Oxford, UK: Elsevier.

Drew, P., & Heritage, J. (1993). *Talk at work: Interaction in institutional settings*. Cambridge, UK: Cambridge University Press.

Firth, A. E. (1995). *The discourse of negotiation: Studies of language in the workplace*. Oxford, UK: Pergamon.

Garcia, A., & Jacobs, J. B. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on Language and Social Interaction, 34*(4), 337-367.

Heritage, J., & Roth, A. (1995). Grammar and institution: Questions and questioning in the broadcast news interview. *Research on Language and Social Interaction, 28*(1), 1-60.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. San Fransisco, CA: Chandler Publishing.

Kärkkäinen, E. (2003). *Epistemic stance in English conversation: A description of its interactional functions, with a focus on i think*. Amsterdam, NL: John Benjamins.

Mason, R. L., Gunst, R. F., & Hess, J. L. (2003). *Statistical design and analysis of experiments: With applications to engineering and science* (2nd ed.): Wiley.

Maynard, D. W. (2003). *Bad news, good news: Conversational order in everyday talk and clinical settings*. Chicago, IL: University of Chicago Press.

O'Neill, J., & Martin, D. (2003). *Text chat in action.* Paper presented at the ACM Conference on Groupware (GROUP 2003), Sanibel Island, FL.

Pomerantz, A., & Fehr, B. J. (1991). Conversation analysis: An approach to the study of social action as sense making practices. In T. A. van Dijk (Ed.), *Discourse as social interaction: Discourse studies, a multidisciplinary introduction, volume 2* (pp. 64-91). London, UK: Sage.

Pomerantz, A. M. (1984). Giving a source or basis: The practice in conversation of telling 'how i know'. *Journal of Pragmatics, 8*(607-625).

Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction*. Thousand Oaks, CA: Sage.

Sacks, H. (1962/1995). *Lectures on conversation*. Oxford, UK: Blackwell.

Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*, 289-327.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*, 1075-1095.

Schegloff, E. A. (1990). On the organization of sequences as a source of 'coherence' in talk-in-interaction. In B. Dorval (Ed.), *Conversational organization and its development* (pp. 51-77). Norwood, NJ: Ablex.

Soller, A., & Lesgold, A. (2003). *A computational approach to analyzing online knowledge sharing interaction.* Paper presented at the 11th International Conference on Artificial Intelligence in Education, AI-ED 2003, Sydney, Australia. Proceedings pp. 253-260: Amsterdam: IOS Press.

Strijbos, J. W. (2004). *The effect of roles on computer supported collaborative learning.* Unpublished Dissertation, Ph. D., Open Universiteit Nederland, Heerlen, the Netherlands.

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.

ten Have, P. (1999). *Doing conversation analysis: A practical guide*. Thousand Oaks, CA: Sage.