Quantitative Analysis of Interaction Patterns in Online Distance Education

Allan Jeong

With increased reliance on student-student interactions in distance education, computer-mediated communication has been the focus of much research over the last few decades. In this area of research, content analysis is commonly used to identify and classify students’ utterances into specific categories, followed by an analysis of the frequencies of utterances observed within each category (Rourke, Anderson, Garrison, & Archer, 2001). However, this approach generates results that are mainly descriptive rather than prescriptive in nature, reporting for example the frequencies of arguments, challenges, and explanations observed in a discussion. Message frequencies alone provide little information to explain or predict how participants respond to given types of messages, how response patterns are influenced by latent variables (for example, message function, content, communication style, response latency) and exogenous variables (such as, gender, personality traits, discussion protocols, type of task), and how particular response patterns improve the quality of discussions and assist groups in achieving the desired learning outcomes (Jeong, 2005).
At the heart of this issue is the question of what to examine and code in online discourse (examples include cognitive, meta-cognitive, social behaviours; individual versus group; message versus sentence units) and how to analyze the discourse data (such as, frequency counts, response probabilities, Markov chains) in ways that provide findings that are meaningful, insightful, and are of predictive and strategic value. A myriad of models and approaches have been developed and used to elucidate, make more explicit, and operationally measure the form, function, and the dynamic and interactive nature of online discourse. As a result, the following sections present brief descriptions of some of the quantitative methods developed and used by researchers to study online communication at a micro-analytic level. Key authors and articles are cited in this section to highlight and illustrate specific methods. These methods include quantitative content analysis (Rourke et al., 2001), social network analysis, sequential analysis (Jeong, 2005), hidden Markov modelling with multidimensional scaling (Soller, 2004), structural equation modelling (Garrison, Cleveland-Innes, & Fun, 2010), and path analysis (Jeong, Lee, & Kim, 2011). The section begins with a review of the quantitative content analysis method on which many if not all the subsequent and more sophisticated analytic methods are based in varying degrees. Immediately following the descriptions and analysis of each analytic method will be a listing of their major limitations along with suggested directions for future research.

QUANTITATIVE CONTENT ANALYSIS (QCA)

QCA is the foundational method on which many if not most other quantitative methods are based. It is used extensively in computer-mediated communication (CMC) research to determine the nature, function, and quality of messages in relation to a specific task or cognitive function such as critical thinking and argumentation. Rourke et al. (2001) describe the procedures in detail and identify some of its methodological challenges and issues. To conduct QCA, researchers: 1) identify representative samples of the communication they wish to study; 2) create a coding scheme and protocol for identifying and classifying each unit of meaning into a specific category and train coders to use this protocol; 3) compare codes between coders to test for inter-rater reliability; and 4) analyze the frequencies of units observed within the categories and/or test for relationships between categories, outcomes,
and other variables to produce either a descriptive or experimental study.

Researchers have used these procedures to examine the latent content (as opposed to the surface-level content such as number of words or misspellings) in order to determine the frequency of utterances that serve a specific social or cognitive function (e.g., make a claim, question the claim). For example, Gunawardena, Lowe, and Anderson’s (1997) interaction analysis model was used to classify student postings in online debates. Their model was designed specifically to capture the social knowledge construction process. It consisted of 21 categories (e.g., statement of opinion, disagreement, clarifying meaning, testing against personal experience, and summarizing of agreements) organized sequentially into five main phases that identify the main stages of the knowledge construction process. After coding the discussion transcripts with their interaction model, they identified and described specific moments in the discussions where students progressed from one phase to the next phase to validate their five-phase model of the knowledge construction process. Overall, this and other proposed interaction models serve as useful tools for measuring and providing quantitative descriptions of the sorts of behaviours (or speech acts) that take place in many online discussions.

One of the main challenges in using QCA, as noted by Rourke et al. (2001), is that students’ online postings often addresses multiple functions. As a result, researchers often struggle in their attempts to establish a reliable and consistent way to parse each posting into meaningful units prior to coding each unit. One single unit of meaning can be found either within a phrase, a sentence, or an entire paragraph. Studies that have used QCA rarely if ever report any measure of inter-rater reliability to establish the extent to which the postings are similarly and consistently parsed into units of meaning. The second challenge is that the more codes that exist within a coding scheme, the more likely the level of inter-rater reliability will decrease. As a result, the process of coding discussion transcripts is often a very time- and resource-intensive task. The following approaches, computer-scripted discussions and auto-coding with machine-based learning, have been used to address some of these issues.

**Computer-Scripted Discussions**

Computer-scripted discussion systems have been designed specifically to scaffold and by default code or tag each student’s postings. Numerous
text-based communication tools have been developed to support, for example, collaborative argumentation by presenting students with various response options/prompts and rules of argumentation within the discussion environment. For example, Loll, Pinkwart, Scheuer, and McLaren (2011) have recently developed a threaded discussion tool called LASAD (Learning to Argue: Generalized Support Across Domains) that helps students to classify the function of their messages (e.g., claim, supporting evidence, rebuttal) prior to posting a message to the discussion. When a message is posted, the category that the student assigns to the message (e.g., argument, challenge, explanation) is explicitly displayed in the message subject heading. In ShadowPD forum (Jonassen & Remidez, 2002), constraints can also be placed on message–response sequences such that messages are attached to responses by a set of constrained links so that, for example, claims can only be linked to supporting evidence, and counter claims can only be linked to rebuttals. The technique of placing constraints on what types of messages can be posted to a discussion, and the use of labels to mark the function of each message, has been applied in other asynchronous discussion environments such as Fle3 (Leinonen Virtanen, & Hakkarainen, 2002), Ntool (Beers, Boshuizen, & Kirschner, 2004), and in live chats such as AcademicTalk (McAlister, Ravenscroft, & Scanlon, 2004).

One advantage of using computer-scripted discussions is that each posting is intended to serve one and only one function at a time. As a result, the unit of meaning or speech act that each student intends to convey/execute within a posting is explicitly identified and classified by the student. Another potential advantage is that the codes that are assigned to each posting are determined by the intentions of the discussion participants, and not by the experimenter. This might suggest that inter-rater reliability is of a lesser concern or issue, but that is not necessarily the case. Jeong and Juong (2007) implemented five message categories (argument, explanation, evidence, critique, other) to support collaborative argumentation and found that students classified their postings only 51% of the time with Cohen’s Kappa = .31 (Cohen, 1960). In contrast, a comparison of two coders’ classifications of the students’ postings using the same coding scheme produced a Cohen’s Kappa of .87. As a result, future research on these computer systems will need to focus attention on testing and reporting the accuracy of students’ codes and finding ways to increase accuracy. In addition to this potential problem with inter-rater reliability is that the discussion protocol in itself
is likely to influence how students interact with one another. As a result, it cannot be determined to what extent the interactions observed within these types of computer systems can be generalized to discussions produced in non-scripted discussion environments.

**Machine-Based Learning Systems**

Machine-based learning systems use computational linguistics to classify online discourse automatically. For example, Rosé et al. (2008) developed a suite of tools called TagHelper that automatically implements a number of different algorithms to segment and classify a student’s utterances into speech acts. Using a combination of strategies that include analysis of text features and the sequential relationship of one speech act to another speech act, TagHelper was able to produce acceptable levels of reliability (ranging from Cohen Kappa values of .60 to .96) in coding discussions across multiple dimensions defined in a coding scheme developed by Weinburger and Fischer (2006). Cohen Kappa of .60 was achieved in coding micro-level argumentation, and .70 for coding macro-level argumentation. See Rosé et al. (2008) for complete details about the various methods and measures of effectiveness.

One benefit of using machine learning systems to code group discussions is that the discussions need not be coded by the experimenter or the students, thus making it possible to code and analyze a larger corpus of data while avoiding the use of discussion protocols and message tagging schemes that might have unintended effects on the way students interact with one another. Furthermore, this approach can be incorporated into a discussion environment, as it has in the ARGUNAUT system (McLaren et al., 2007), to analyze online discourse automatically in real-time to help instructors moderate discussions more effectively. One of the requirements of using machine-based learning is that the experimenter must manually code an initial corpus of data to provide data that can be used to train the system. Furthermore, this process must be repeated when analyzing different types of discourse using different coding schemes that address different instructional goals and task demands.

Regardless of what methods are used to code student discourse in online environments, the QCA method of classifying and observing discourse move frequencies is limited in its ability to identify stable and meaningful patterns
in student behaviours—patterns that can be generalized across different student groups, discussion topics, task structures/demands, and domains. By relying simply on observed frequencies, one study might in theory find that one group posted a significantly larger number of questions but significantly fewer explanations than another group. Or, the study finds that the first group posted proportionately more questions than explanations than the second group. Examining these types of patterns might shed some light, for example, on how a particular intervention helped to encourage more questions from students. However, there is no basis on which to establish what proportion of questions-to-explanations is to be considered an acceptable level and to be established as the norm. Furthermore, the observed frequencies do not help to explain the immediate context and discourse moves that elicit students’ questions or to determine the extent to which students’ questions elicit explanatory responses. In other words, simple frequencies do not provide insights into the sequential relationships between dialog moves to fully capture the action–reaction dynamics between discussion participants. To examine the relationships between discourse moves and discussion participants, and in order to build on the frequency counts produced from using QCA, researchers are using the methods of social network analysis, Markov chain analysis, and sequential analysis.

**Social Network Analysis (SNA)**

This method examines interactions between participants by producing quantitative measures that are conveyed visually via network graphs or sociograms. Coloured nodes in the graphs represent individual participants or a subset of participants. The edges that link the nodes identify participants who produced at least one response to the messages of another participant (out-degree values). Alternatively, the edges can also be used to identify individual participants who received at least one or more responses from certain participants (in-degree values). The distance between the nodes conveys how often one participant responded to or received responses from a certain participant. The shorter the distance between two participants, the greater number of responses exchanged between the two participants. When using SNA to analyze the observed frequency of exchanges between individual participants, one can, for example, measure density (how often
the participants overall respond to one another’s postings) and centrality (to what extent certain discussants play a central role across multiple conversational threads (Scott, 2000). As a result, *density* describes the general level of cohesion between the participants, and *centralization* describes the extent to which this cohesion is organized around particular participants.

Using the SNA method, de Laat, Lally, & Lipponen (2007) conducted a study to determine how interaction patterns between students in a collaborative project changed over time. In figure 15.1 are three sociograms produced from the analysis of out-degree values (number of times a student posted responses to certain students) observed at the beginning, middle, and final phases of the group project. The findings revealed that group cohesion in the middle remained mostly the same while out-degree centralization went up. While decreases occurred in both level of cohesion and centrality near the end of the project, certain members in the group continued to communicate actively with most if not all of the other group members. Students were interviewed (using the critical event recall method) to identify the factors that contributed to these changes in interaction patterns (e.g., socializing and group norming at the beginning, breaking into small work groups, taking on the role of group moderator, and so forth).

Overall, this case study demonstrates that SNA can be used as a descriptive tool to identify interaction patterns between certain students and reveal how interactions patterns change over time. SNA can then be used in combination with other methods to determine the underlying factors (e.g., what, why and how students are communicating with other students) that contribute to observed changes in interaction patterns and whether certain changes in interaction patterns lead to better group learning and group performance. A limitation of using SNA in this manner is that it remains to be seen if group cohesion and centrality is a reliable predictor of group performance and learning given the various ways in which groups structure and coordinate tasks over the course of a group project. In addition, SNA graphs only reveal information on who is interacting with whom and not on the nature and function of the interactions that take place between participants. As a result, research can be conducted to see if predictive validity can be improved by comparing graphs that convey the relationships between students within a subset of exchanges, such as exchanges with opposing viewpoints (claim→disagree, claim→counter-evidence) versus exchanges with supporting viewpoints (claim→agree, claim→supporting evidence).
Figure 15.1 Change in group interaction patterns in a collaborative group project (de Laat, Lally, & Lipponen, 2007).
MARKOV CHAIN ANALYSIS

To examine the functional relationships between messages and responses, attempts have been made to identify patterns in the relationships between messages (Levin, Kim, & Riel, 1990; Newman, Webb, & Cochrane 1995; Gunawardena, Lowe, & Anderson 1997; Sudweeks & Simoff 1999; Fahy, Crawford, & Ally 2001). Levin, Kim, and Reil (1990) attempted to map and analyze message flow. Sudweeks and Simoff (1999) applied neural network analysis by assigning numerical values to the strength of interrelations between messages. Gunawardena, Lowe, and Anderson examined transitions between phases of critical thinking to illustrate the social construction of knowledge. All of these studies however fall short of providing a robust, more precise, and process-oriented method to measure and visualize student interactions in ways that can enable researchers to determine how specific dialog sequences trigger deeper discussions, cognitive processing, and learning.

Given the complexity and dynamic nature of discourse, dialog-move sequences do not always unfold in orderly and predictable ways. Soller (2004) believed that this is a reason the simple frequencies of each dialog
move performed by learners did not distinguish learners who scored high versus low on a post-test measuring knowledge acquisition. As a result, Soller incorporated a process-oriented approach that examined how interactions unfold over time by producing transitional state diagrams to convey how likely (or the probability) one dialog move was followed by another dialog move (e.g., inform, acknowledge, request information, discuss with doubt, agree). This interaction data (sometimes referred to as Markov chains), combined with post-test scores, were analyzed using multidimensional scaling to reveal clusters of three- to four-event chains that were observed among high performing groups (for example, request info → explain → agree; request info → explain → request clarification → provide clarification) and low performing groups (such as, propose → explain → acknowledge; propose → express doubt).

This particular application of Markov chain analysis produced findings to reveal two, three, and four dialog-move sequences that were associated with and were believed to help students achieve superior learning. These findings reveal the types of interactions to be encouraged and discouraged either by the instructor or by discourse systems that incorporate machine learning and natural language processors for automated gauging and monitoring of student discourse. Further understanding as to how these longer chains of dialog moves develop requires an even closer micro-genetic examination of the transitional probabilities between dialog-move pairs and the factors that positively and adversely affect the probabilities that result in improvements and breakdowns in the group communication and group learning.

**SEQUENTIAL ANALYSIS**

To conduct a finer-grain micro-analysis of the transitional probabilities between specific dialog moves, Jeong (2006) used sequential analysis to determine: a) how the use of conversational language (e.g., making references to participants by name, saying thank you, and use of greetings and emoticons) affected the probabilities of certain responses elicited by arguments, challenges, explanations, and presentation of supporting evidence; and b) to what extent the observed probabilities are significantly higher and lower than the expected probabilities based on z-scores (Bakeman &
Gottman, 1997). The findings revealed that the (argument→challenge→explanation) interaction was more likely to emerge from students’ interactions when students used more conversational language when presenting arguments, challenges, and explanations.

Figure 15.2 Response patterns produced from messages with or without conversational language.

ARG = argument, BUT = challenge, EVID = supporting evidence, EXPL = explanation, c denotes messages presented in a conversational style, + denotes transitional probabilities significantly higher than expected based on z-scores at p < .01; the size of glow surrounding each dialog move conveys the number of times the dialog move was observed; first digit in node conveys number of times dialog move was observed, and second digit following the > symbol is the total number of messages posted in response to the dialog move. State diagram produced with DAT software (Jeong, 2011).

For example, figure 15.2 shows that arguments presented without conversational language elicited challenges in 52% of responses, compared to 90% when arguments were presented with conversational language. In addition, challenges presented without conversational language elicited explanations.
in only 9% of responses compared to 23% when presented with conversational language.

Although similar analysis can be conducted with SNA software by replacing the nodes with dialog moves (instead of the names of individual participants), figure 15.2 demonstrates that comparing and identifying differences in patterns between groups can be conducted more effectively by: 1) keeping the positions of each dialog move identical in both diagrams; and 2) varying the thickness (instead of the length) of the edges connecting dialog moves in direct proportion to the observed transitional probabilities between dialog moves. Further clarity can be achieved by varying the saturation of the edges (e.g., solid black or light gray edges) in relation to the observed probabilities or to z-scores that determine whether the observed probabilities are significantly higher or lower than the expected probabilities. To identify patterns that convey how dialog moves emanate from prior dialog moves in order to provide a historical perspective, historical state diagrams can be produced (Jeong, 2011) to convey how likely each dialog move elicited the dialog move of interest. Overall, sequential analysis and these particular methods for increasing precision in pattern identification should enable future studies to: 1) determine to what extent differences in discourse patterns/processes (particularly patterns among message–response pairs or first order chains as opposed to longer higher order chains) account for variance in group performance and learning outcomes; and 2) better predict how particular dialog moves under certain conditions influence response behaviours in ways that help to produce dialog move-sequences/chains that lead to significant gains in group performance and learning.

### Structural Equation Modelling and Path Analysis

Structural equation modelling and path analysis are two other methods used to examine the dynamic and emergent nature of interactions between participants. Using structural equation modelling (SEM), Chen and Chiu (2008) examined how earlier messages affected later messages along five dimensions: (1) evaluations (agreement, disagreement, or unresponsive actions); (2) knowledge content (contribution, repetition, or null content); (3) social cues (positive and negative acknowledgments); (4) personal information; and (5) elicitation (eliciting response or not). By analyzing 131
messages across seven topics in a university mathematics discussion forum, this study generated a SEM model that conveyed the causal relationships between these five categories of messages. The study found that: a) a disagreement or contribution in the previous message increased the likelihoods of disagreements and social cue displays in the current message; and b) online discussion messages that disagreed with an earlier message were more likely to elicit responses. Like the findings generated with sequential analysis (Jeong, 2006) presented above, Chen and Chiu’s findings suggest that instructors can monitor online discussions at the message level to promote critical thinking, facilitate discussion of controversial topics, and reduce status effects.

Jeong and Lee (2010) used path analysis (a variation of SEM where only one indicator is needed to measure each variable/behaviour) to determine how five particular online behaviours are directly and indirectly related to the quality of students’ postings in online debates—five behaviours that online instructors might use to set minimum participation requirements. The five behaviours were: 1) messages posted to initiate a new discussion thread; 2) different days in which the student made one or more postings; 3) messages posted in reply to another student’s posting; 4) replies elicited from each student’s posting; and 5) reciprocated replies (or uptakes) posted by each student. The path analysis produced a model that suggested that: a) requiring students to post a certain number of replies to other students’ postings could have an adverse effect on the quality of students’ postings; and b) instructors can set requirements on number of opening arguments/threads posted and number of different posting days to increase directly the number of responses elicited by each student’s messages and number of reciprocal replies in order to increase the quality of students’ postings.

Both SEM and path analysis provide useful tools to determine the possible direct and indirect cause–effect relationships between particular student behaviours and outcomes in online discussions. Using these two methods to distinguish behaviours that have direct from indirect effect on any given target behaviour enables researchers to determine strategically which one or two behaviours online instructors can key on, monitor, and promote to achieve the desired target behaviours. By identifying and promoting just a few key behaviours, online instructors can avoid imposing on students with too many posting requirements and still elicit the target behaviours. At the same time, one of the limitations of these two methods
is that they cannot be used to test the causal direction between behaviours. The direction of the arrows in a structural equation or path analysis model represents only the researcher’s hypotheses as to how one behaviour affects another. Furthermore, there is always the possibility that several alternative models also fit the data equally well. Nevertheless, these two approaches are effective means to improving our structural understanding of the causal relationships within a complex system of behaviours observed in online discussions.

**IMPLICATIONS FOR FUTURE RESEARCH**

The methods presented above provide just a sample of the quantitative methods that researchers have developed and used to achieve a more precise and in-depth understanding of discourse in online environments. To refine these methods and to establish the validity of the findings produced with these methods, further research is needed to determine what interaction models and typologies will produce the most useful findings when used to conduct a micro-genetic analysis of online discourse. To achieve this, researchers will need to develop and articulate more precise models and theories of collaborative knowledge construction (across different task-demand structures). The detailed model of collaborative knowledge construction articulated by Stahl (2004), for example, can be used to frame the identification, selection, and operation of collaborative interactions to build interaction models that are coherent, complete, and conceptually sound. At this time, a large number of interaction models exist that are similar in many ways while possessing their own set of nuances and idiosyncrasies (Marra, Moore, & Klimczak, 2004). Establishing a theoretical framework will help researchers synthesize and integrate existing interaction models and use these models to conduct more systematic body of research.

In addition, the interaction models that have been proposed by researchers differ widely in the dimensions of group interaction represented in the models that serve as the focal points of the analysis. For example, Henri’s (1992) model of interactivity consists of three categories (explicit interaction, implicit interaction, and independent statement) that identify structural relationships in terms of how a student’s message is related to the previous messages of other students. Rourke’s (2001) social presence model consists
of three categories in which its first category (interactive) is also structural in nature. This category is then combined with the cohesive category that addresses processes related to group building/processing, and the affective category that addresses the emotional dimension of group discourse. In contrast, Gunawardena, Lowe, and Anderson’s (1997) model consists of five main categories that represent solely the cognitive operations that group members must perform to construct shared meaning and knowledge. Studies that use these models often examine each dimension in isolation, and those that do examine the inter-relationships between dimensions often report descriptive findings that are of little or no strategic value.

To conduct the research needed to help achieve a full understanding of how discourse leads to improvements in group learning/performance, particular attention must be focussed foremost on the cognitive operations exhibited in dialog move and move sequences. The assumption here is that the cognitive processes that learners perform is the primary determinant of student learning. With the cognitive dimension as the central focus of the discourse analysis, researchers can systematically examine how changes across other dimensions (e.g., social, emotional, meta-cognitive) affect changes in discourse processes. For example, Garrison, Cleveland-Innes, & Fung (2010) used structural equation analysis to reveal the extent to which learners’ social interactions and interactions with instructors impacted the cognitive processes performed by learners. Furthermore, Jeong (2006) used sequential analysis to examine how conversational language (a social dimension) affected positive changes in discourse patterns observed in online group debates.

Finally, future research is needed to determine if and to what extent the integration of existing interaction models is even possible or desirable. Researchers who use qualitative methods to study online discourse often argue that each learning community possesses its own unique set of practices that reflect and are situated within a social-cultural context. Each learning community’s set of practices thereby shapes and constrains the discourse and discourse process. As a result, it may not be theoretically possible or even desirable to develop interaction models that can be generalized across multiple contexts.

Future research can be conducted by applying the quantitative methods described above to determine: a) to what extent interaction models can be developed and applied across contexts; b) if such models only work when
the typologies articulate the discourse processes at the macro- versus the micro-level; and c) if differences in discourse between different learning communities stem from differences in dialog move typologies, or differences in dialog sequences/processes. All three of these issues can be addressed by examining which models and approaches reveal discourse patterns that best predict learning and performance. The extent to which researchers are able to develop and disseminate software tools for building interaction models, classifying and micro-analyzing discourse, and conveying the findings to other researchers and practitioners will likely determine future success in addressing these fundamental questions.

REFERENCES


