

**van Aalst, J. (2012). Commentary: Combining content analysis and social network analysis. *Interdisciplinary Journal of Technology, Culture, and Education*, 7 (2), 84-94.**

The three studies in this special issue explore new approaches for analyzing interaction and learning in virtual communities and blended communities, which involve interaction in both offline and online settings. All three studies combine content analysis informed by sociocultural theories with methods inspired by social network analysis. This commentary attempts to summarize what has been accomplished and explores and the potential for developing techniques.

### Background

Early research on virtual communities was strongly influenced by Chi's (1997) suggestion to quantify the content of discussions by segmenting them into idea units, applying content analysis to each idea unit by applying a code to it, and finally analyzing the code frequencies statistically. Over the years, a variety of coding schemes have been developed for this approach (e.g. Gunawardena, Lowe, & Anderson, 1997; Henri, 1992; Meier, Spada, & Rummel, 2007). The coding schemes generally contain few categories, one of which is applied to a specific idea unit; inter-rater reliabilities are high (Cohen kappas > 0.8). Studies that have employed this approach have led to important insights into global features of online discourse, for example they have elucidated the importance of focusing the discourse on developing explanations rather than only answering fact-seeking questions (Hakkarainen, Lipponen, & Järvelä, 2002; Lipponen, 2000). However, over time the limitations of this approach became apparent. One problem was that these analyses took ideas out of the context in which they occurred. For example, in commenting on the analysis of Hakkarainen et al.

(2002), Stahl (2002) pointed out that the analysis ignored the fact that a statement of an idea could be a response to a question or a modification of an earlier idea. It was as if ideas were transferred directly from students' heads to the discussion forum. As the field of computer-supported collaborative learning (CSCL) matured it became clear it would be concerned with the analysis of intersubjective processes of meaning making, that is, on how learning is accomplished via interactions in a CSCL environment (Suthers, 2006).

Although situated cognition theories (Brown, Collins, & Duguid, 1989; Hutchins, 1995) have influenced CSCL research, the field remains somewhat fragmented by camps using sociocognitive and sociocultural theories. One option regarding the latter is provided by sociocultural theories of meaning derived from Wittgenstein's philosophy of language and literary criticism has become increasingly influential in educational research. According to Wittgenstein a word does not have an objective meaning that can be read fully from a dictionary but receives meaning anew each time it is invoked in discourse. Bakhtin (1981) argued that language is inherently dialogic, that an utterance is always anticipated by earlier utterances and anticipates future ones. Another option is Socio-Historical Activity Theory (AT; Cole & Engeström, 1993), which emphasizes the mediational role of tools and artifacts in activity. The three studies in this special issue all draw on these theories in different ways.

All three studies are also influenced by Social Network Analysis (SNA; de Laat, Lally, & Lipponen, 2007; Haythornthwaite, 2002). SNA is an application of graph-theoretical methods, and is used for analyzing the social structure of offline and online communities; it analyzes networks in which participants are represented

by nodes and information flow from one participant to another by links. It can be used to analyze the quantity of interaction, but it cannot answer questions about the quality of the information communicated and the interaction. Moreover, just as words need to be understood in sociocultural context so must, as Impedovo, Ligorio, and Law (this volume) review, identity. Which aspect of a persons' identity should be used to interpret SNA results? Nevertheless, one can in principle attempt to construct networks in which the entities represent more generalizable constructs, and use mathematical techniques also used in SNA to analyze the properties of those networks. Two of the articles attempt this (Annese & Traetta, this volume; Impedovo et al., this volume).

#### Positioning Network Analysis (PNA)

Annese and Traetta provide a rich theoretical analysis of identity (i.e., Self) that draws on Dialogical Self Theory and Positioning Theory. Dialogical Self Theory, developed by Hermans, is primarily based on Bakhtin's concept of identity as storytelling, according to which the storyteller uses multiple voices that refer to a polyphony of selves, that are in dialogue with each other. As Annese and Traetta explain, "Dialogic Self Theory is a multiplicity of I-positions; each I-position provides the self with a voice, and these multiple voices draw up and oppose each other dialogically, producing different positionings organized in a coherent identity plot" (p. 4). Thus, just as words do not have context-independent meaning, so identity is sociohistorical; across specific situations different voices in a specific person's identity may dominate over others. The concept of *positioning* is central in the Hermans' theory; the authors provide a broad exploration of its interpretations as well as methods for investigating

them. A position can be considered a way of being in a specific situation; the authors settle on Hermans' Personal Position Repertoire as a starting point for their own approach. One of the modifications is to extend Hermans' approach to collective positionings.

In developing their method, Annese and Traetta first elaborate a comprehensive grid covering individual, collective, interpersonal, intergroup, and boundary positionings. They then code each computer note using this grid to identify how participants position themselves toward the community, thus marking their degree of involvement (p. 11). In the second stage of coding the authors identified links between the positionings, distinguishing between elicited and eliciting positionings. These procedures resulted in an adjacency matrix of elicited and eliciting positionings, which the authors analyzed using the methods of SNA.

The interpretation of the networks seems relatively straightforward. The nodes do not represent the participants as static identities, but rather the positionings of all the participants obtained from coding the discourse. The greater the network density, the more of the potential eliciting or elicited relations between the positionings exist in the discourse. A positioning with a large degree centrality is then more involved in eliciting or being elicited by other positionings. A network with a density close to one and little variation in degree centrality would represent a uniform identity in which nearly all the positionings are involved similarly; a network with lower density and more variation in degree centrality would represent an identity in which some positionings are only marginally involved, and thus would have more structure. PNA may become

an important alternative to SNA because the nodes are defined more precisely as theoretical constructs, whereas in SNA they refer to participants very generally; the links represent the strengths of association between the nodes, although it would be useful to explore additional types of association besides elicitation. And while SNA is only a quantitative analysis, PNA combines qualitative and quantitative analyses; although the qualitative analysis still takes utterances out of their context, some of that context is put back into the analysis by representing the associations between the positionings. The authors further deepen the results by examining the construct identity at multiple levels of social organization (individual, interpersonal, and collective), an important issue for CSCL (De Wever, Van Keer, Schellens, & Valcke, 2007; Suthers, 2006).

Annese and Traetta applied PNA to the online and offline discourses, on the same topic, by one collaborative group; this section of the article provides a very clear illustration of what PNA can reveal about the nature of identity in collaborative learning contexts. Not surprisingly, they found the Internal/Individual positioning had the greatest centrality in their application of PNA to discourse data from online and offline discourse by the same collaborative group. This is what we would expect from a traditional analysis of identity. However, many other positionings also were important, such as Internal/Collective, Boundary, and Direct/Interpersonal (for definitions see Annese & Traetta, Table 1). The authors explain: "In both settings, the otherness of social positionings ... is vital to the construction of the self. The sense of community is fundamental in the construction of individual identity." (p. 20)

#### Reformulation Analysis

Traetta, Annese, and Loperfido ( this volume) also start from a dialogical perspective inspired by Bakhtin, and review that meaning is not inherent in a message but in “the position it occupies between speakers” (p. 3). As these authors point out, “Bakhtin’s {1981) concept of ‘responsive understanding’ clearly refers to the unavoidable presence of an interlocutor in our speeches, which are always addressed to someone even if they are physically absent.” However, whereas (Bakhtin) characterized dialogism as a feature of language, these authors, following Marková, consider it an epistemological position. The method of Traetta and colleagues focuses on the concept of *reformulation*, which is more than simply a “textual repetition of other’s discourse” but involves a “dialogical revision of the evocated voices” and “tries to give a new sense to the recalled discourse” (p. 5). Specifically, these authors focus attention to one kind of reformulation: *not-paraphrastic* reformulation, which involves considerable distance from the text that is being reformulated; they claim these reformulations can be identified from meta-discursive clauses such as “as you said.”

Much of this article is taken up by discourse analysis to show how the concept of reformulation can be applied to online and offline utterances. This presentation ends with a 2 × 2 framework that distinguishes between discursive and cognitive reformulations, and between individual and collective formulations; the authors also distinguish between contextual reformulations, in which the reformulation is in the same medium as the source, and blended reformulations, in which it is not. The framework is simple through its parsimony, and therefore suitable for analysis, but the naming of some of the categories leaves something to be desired. For example, individual reformulations are signaled by “clauses” but collective

reformulations by “comments;” the former refers to a grammatical structure but the latter to a communicative function. The label “contextual” also is confusing because all speech is contextual according to Wittgenstein’s philosophy of language. Perhaps it would be better to consider discourse more holistically, as one that involves multiple modalities including online writing, offline talk, gesturing, the use of inscriptions, and so forth.

Nevertheless the results from this study clearly reveal the potential of the method. The authors use the framework to code both the online and offline discussions occurring in the same course of study, identifying 1266 reformulations. It is a nice feature of this study that it analyzes the online and offline aspects of the participant’s discourse together; relatively few studies do so; and Annese and Traetta analyze online and offline discussions separately. However, Traetta and colleagues present their results according to the medium in which the reformulation occurred.

They found substantial differences between the two media. In the reformulations in the online medium (Traetta et al., , Table 8), the vast majority (94.23%) of reformulations referred to sources in the same medium, and most (68.08%) were individual discursive reformulations. These results suggest that a sense of community in the online environment was not highly developed: Relatively few reformulations were made on behalf of a group. In addition, there were relatively few cognitive reformulations (9.22%). By contrast, the reformulations made in the offline medium (Traetta et al., , Table 9) referred to sources in the other medium more often, and included more collective and cognitive reformulations (22.47% for the last item). These results suggest that there may have been a

greater sense of community in the offline environment and that that online work made greater impact on the offline discussions than the converse. The greater evidence for (meta-) cognitive reformulations in the offline discussions is surprising because (asynchronous) online discourse is generally considered more reflective than offline discourse (Lin, Hmelo, Kinzer, & Secules, 1999).

Though the findings that Traetta and colleagues report are potentially informative, the use of their method needs to be part of a multi-method approach. The present study involved comparisons of reformulations in offline and online media, but no data were presented that could be used to interpret the observed differences. For example, how many participants were involved in the offline discussions? How many responded to online posts or were aware of them through reading? The authors carried out SNA and claim that the reported results of the reformulation analysis “produced a specific and dynamic evaluation of the SNA representation” (p. 14) but unfortunately did not report the SNA results.

#### Analysis of Inter-action

Rather than starting from Bakhtin, Impedovo and colleagues use socio-historical Activity Theory (AT) and Speech Analysis Theory (SAT). Similar to Annese and Traetta they use methods from SNA; the nodes in the networks represent the main elements of AT rather than positionings or participants.

These authors segment each computer note into speech actions, each one of which has a distinct communicative goal; they then code each segment using AT the elements as categories, replacing “division of labor” by “interaction” to capture the different forms of interaction involved in the activity. Then each



category in the AT is assigned subcategories for showing how the category may appear in the online discussions. For example, the AT category Subject includes the subcategories self-reference, cognitive elements, internal identity, external identity, embodiment, role, and belonging (Impedovo et al., Appendix A). The coding process is described in detail, with a clear description and example of each subcode. Overall, this is a rather fine-grained and labor-intensive analysis. However, only the frequencies for the AT categories and the most frequently coded subcategory for each AT category are reported (Impedovo et al., Table 3); it would have been nice to see more of how the subcategories are populated. The results include some interesting findings: The most frequently coded AT categories are Interaction (28.6%) and Object (23.1%); there were few code instances for “rules and community values” and “outcome.”

In the second step of the analysis, the authors used a similar procedure to that of Annese and Traetta to create an adjacency table: For each communicative action they looked for *connected actions* that could be considered as eliciting a new action or being elicited by another action (Impedovo et al., Table 4). However, my interpretation of the results differs somewhat from that of the authors. First, *all* of the segments were coded as being elicited; this is surprising because in an extended discussion there should be new communicative actions that are not evidently related to previous actions—a participant may raise an idea because no one has yet done so. Second, the majority (68.9%) of segments did not elicit any communicative action; thus these did not have a direct effect on the development of the discourse. To further facilitate interpretation of these results it would be interesting to know how the tutor was involved in the discussion; for example if the tutor posted a few questions to which all students responded

independently, it would explain why so few segments elicited communicative actions. Zhang, Scardamalia, Reeve, and Messina (2009) have used SNA to show how social configurations in which the teacher had different roles produced different results in terms of knowledge diffusion and knowledge quality. Information about reading patterns would also be helpful; as Wise, Speer, Marbouti, and Hsiao (2012) have shown, reading—or “listening”—behaviors can account for the majority of time students spend in online discussions.

The final step in the analysis uses SNA methods; it seems an interesting way to characterize the data (Impedovo et al., Figures 2 and 3). The resulting network is similar to the diagram usually employed in studies based on AT, except that the distances between the AT categories now have metrical meaning; the density of the network is 0.61 and the AT category Rule is only connected weakly to the activity system. However, the results are difficult to interpret. In SNA and PNA the nodes can be said to represent members of populations—i.e., all the participants in a community or all conceivable positionings—so that in sufficiently long discourses we can expect to see evidence of all relations among them. In contrast, the AT categories do not appear to have this relationship to one another. To what extent is a metrical representation of the activity system consistent with the theoretical constructs of activity theory? The centrality plot (Impedovo et al., Figure 3) suggests that Interaction is the AT category most involved in eliciting and being communicated by communicative actions.

### Conclusion

The papers in this special issue provide a valuable attempt to develop new methods for analyzing discourse in blended communities. They are to be noted

for the extent to which they build on sociocultural theories of meaning and identity and for their attempt to develop SNA-like quantitative methods. Through all studies involve coding of computer notes or segments out of the context in which they occur, the content analysis is deeply conceptual, and the analysis of relations between the codes via eliciting or elicited segments to some extent restores the contextuality of the coded segments. In this respect, these studies suggest an advance over earlier methods.

Future research should address at least two issues. One is that more theoretical analysis is needed to evaluate the interpretability of the SNA-like networks that result from analyses of relations between codes. It is technically possible to create SNA-like networks from coding, but the networks also must be theoretically meaningful—even if they provide cogent descriptions of the data. Second, if the strategy turns out to be fruitful, it is a labor-intensive one. It is therefore important to explore to what extent automatic coding can help. (Traetta et al.) used clauses and comments to reformulations, so it may be possible to apply techniques similar to those used for automatic coding of argumentation.

### References

- Annese, S., & Traetta, M. (2013). The dialogical self between the virtual and real in Positioning Network Analysis. *Qwerty*.
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. Austin, TX: university of Texas Press.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences*, 6, 271-315.
- Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (1-46). New York, NY: Cambridge University Press.

- de Laat, M., Lally, V., & Lipponen, L. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for social network analysis. *International Journal of Computer-Supported Collaborative Learning*, 2, 87-103.
- De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Learning and Instruction*, 17, 436-447.
- Gunawardena, L., Lowe, C., & Anderson, T. (1997). Interaction analysis of a global on-line debate and the development of a constructivist interaction analysis model for computer conferencing. *Journal of Educational Computing Research*, 17, 395-429.
- Hakkarainen, K., Lipponen, L., & Järvelä, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In T. Koschmann, R. Hall & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 129-155). Mahwah, NJ: Lawrence Erlbaum Associates.
- Haythornthwaite, C. (2002). Building social networks via computer networks: Creating and sustaining distributed learning communities. In K. A. Renniger & W. Shumar (Eds.), *Building virtual communities: Learning and change in cyberspace* (pp. 159-190). New York, NY: Cambridge University Press.
- Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing* (pp. 117-136). Heidelberg: Springer.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Impedovo, M. A., Ligorio, M. B., & Law, E. H. F. (2013). A method for the analysis of inter-action in online learning. *Qwerty*.
- Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, 47, 43-62.
- Lipponen, L. (2000). Towards knowledge building: from facts to explanations in primary students' computer mediated discourse. *Learning Environments Research*, 3, 179-199.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2, 63-86.
- Stahl, G. (2002). Rediscovering CSCL. In T. Koschmann, R. Hall & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 169-181). Mahwah, NJ: Lawrence Erlbaum Associates.
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1, 315-337.
- Traetta, M., Annese, S., & Loperfido, F. D. (2013). Group dynamics in virtual communities: Reformulation process as a dialogical device. *Qwerty*.
- Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y. T. (2012). Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. *Instructional Science*(Online first), 1-21. doi: 10.1007/s11251-012-9230-9.

Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *The Journal of the Learning Sciences, 18*, 7-44.