Visualizing Dynamic Topic Analysis

Susan C. Herring
Indiana University
1320 E. 10th St., LI 011
Bloomington, IN 47405 USA
herring@indiana.edu

Andrew J. Kurtz
Indiana University
1320 E. 10th St., LI 011
Bloomington, IN 47405 USA
ajkurtz@indiana.edu

CHI 2006, April 22–27, 2006, Montreal, Canada.
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract
In this paper we describe a technique for analyzing textual conversations, Dynamic Topic Analysis, and a tool, VisualDTA, for automatically creating visualizations of data coded according to this technique.

Keywords
Computer-Mediated Communication, Conversation Analysis, Interaction Design, Topic, Visualization

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Within a conversation, even when there are only two participants, the discussion topic changes over time. It is interesting to analyze the evolution of topics in different modes of online conversation—such as IRC, IM, and message board conversations—to observe how they are similar or different from one another and from conversation in other media. Herring [3] has created a technique for analyzing the coherence of conversation in text-based computer-mediated discourse. This technique, Dynamic Topic Analysis (DTA), provides a means to visualize and quantify the structure of topic flow within a conversation. A coding scheme is used to quantify the topic drift, and a visualization based on the coding may be created to display the flow of the topics.

VisualDTA

This research developed an interactive Java application, called VisualDTA, which is used to automatically create a visualization from a DTA coding [6]. In addition to displaying the visualization, VisualDTA provides interactive tools to enhance the topic analysis process. This research was conducted by Andrew Kurtz and Susan Herring and is based on the visualization described in Herring [3].

Utilizing VisualDTA to analyze a conversation involves creating a DTA coding. Typically the coding is entered in Microsoft Excel and exported to a TAB delimited file. The coding file is loaded into VisualDTA and the visualization is created and displayed. The display is a tree with the root at the top and the branches flowing down and to the right. The passage of time in the conversation is represented descending along the y-axis. Rightward movement on the x-axis represents the semantic distance of how topically distant the proposition is from the previous proposition.

The propositions are represented by a letter indicating the relationship between the current proposition and the proposition to which it is responding, called the move (T for "on-topic", P for "parallel shift", E for "explanation", M for "metatalk", and B for "break"; cf. [5]). A proposition that replies to a previous proposition is connected to that proposition using a line. A dotted line is used when the connection is tenuous. A "break" is not connected to any previous proposition. The visualization of a segment of a multi-participant Internet Relay Chat conversation using VisualDTA is shown in figure 1.

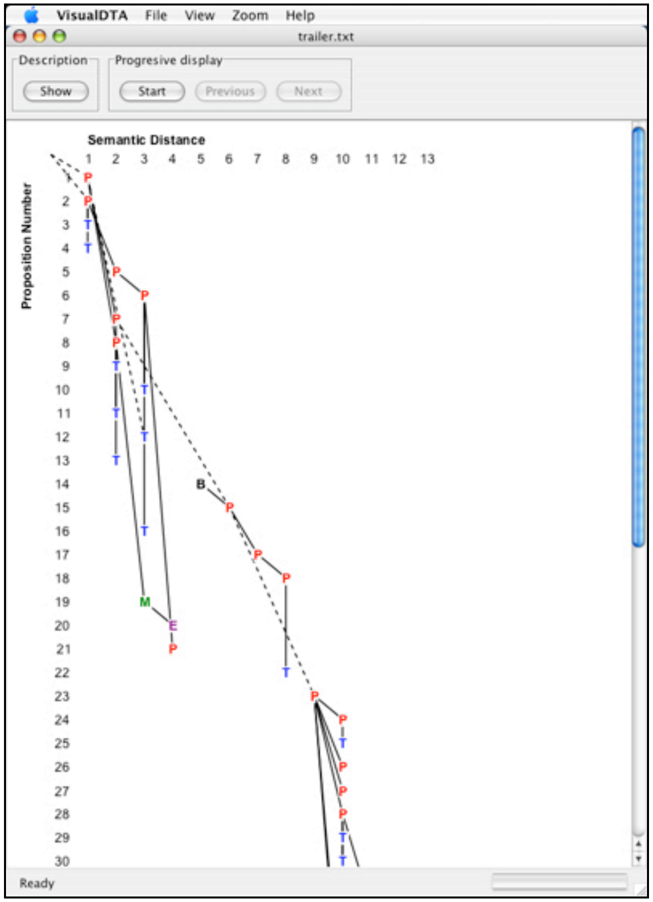

figure 1. VisualDTA displaying an IRC conversation.

When the visualization is displayed, a number of interactive options are available to enhance analysis that may be performed using the DTA technique.

- The text of each proposition can be displayed.

- The visualization can be displayed proposition-by-proposition as the conversation unfolds.
- Hovering over a proposition with the mouse will display information about that proposition, including any attributes coded such as speaker, gender, and role.
- Clicking on a proposition highlights all of the propositions that match an attribute for the selected proposition. For example, all the contributions typed by a specific participant can be highlighted.
- Basic statistics (see figure 3) can be generated, including the average semantic distance and counts for proposition attributes, such as move type, speaker, gender, and role in the conversation.
- Customized statistics can be generated to relate two attributes, such as move type by gender.

Figure 2 shows the use of an interactive feature to highlight the contributions of a specific participant in a dyadic Instant Messaging (IM) conversation.

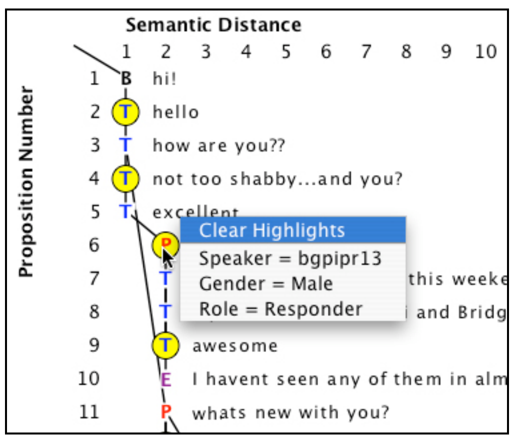

figure 2. Using the interactive options to highlight a specific participant.

Proposition counts

Number of propositions: 76
Avg. semantic distance (all): 0.5
Avg. semantic distance (P): 1.273

Move counts

T: 46 (61%)
P: 22 (29%)
B: 2 (03%)
M: 1 (01%)
E: 5 (07%)

Speaker counts

Speaker	#	Percent
BrookRashell	42	55%
bgpipr13	34	45%

Gender counts

Gender	#	Percent
Female	42	55%
Male	34	45%

Role counts

Role	#	Percent
Initiator	42	55%
Responder	34	45%

figure 3. Basic statistics for the conversation in figure 2.

Usage

Dynamic Topic Analysis has been used to analyze and visualize various types of textual, especially synchronous, computer-mediated communication (CMC). Herring and Nix [4] employed it to compare multi-participant Internet Relay Chat in social and educational contexts. That analysis revealed different patterns of coherence according to the purpose of communication and the presence of a discussion leader. Stromer-Galley and Martinson [11] adapted VisualDTA to analyze web chat about politics, auto racing, entertainment, and cancer support, finding the political chat to be the most coherent.

Recently, we have been using VisualDTA to analyze Instant Messaging exchanges. Text-based IM resembles prototypical casual conversation: It is synchronous, dyadic, and frequently engaged in for social purposes. This research adapts principles of conversation analysis [8] and exchange structure analysis [2, 9] to Instant Messaging.

The data are a corpus of 200 IM conversations produced by undergraduate students at Indiana University during the summer of 2003. The corpus includes male and female same-sex conversations between friends, and cross-sex conversations between strangers in the same peer group. We used VisualDTA to create interactive visual representations of the main sequential patterns that emerged from these data.

Recurrent structures are evident in the VisualDTA representations. Three conversational schemas emerge that are visually distinct, according to the function of the communication: phatic chat, planning an activity, and getting acquainted, the latter of which occurs mainly in cross-sex conversations. Thus VisualDTA not only shows cohesion (interactional meaning); it can also distinguish discourse activities (socio-pragmatic meaning).

A current goal is to identify abstract units of conversation that emerge using VisualDTA: meaningful chunks of structure, such as 'greeting sequence' (figure 4), 'topic drift' (figure 5), and 'digression' (figure 6), that combine to form longer conversations.

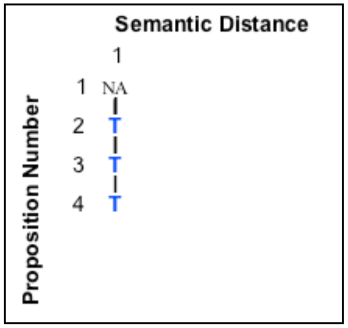

figure 4. Greeting sequence.

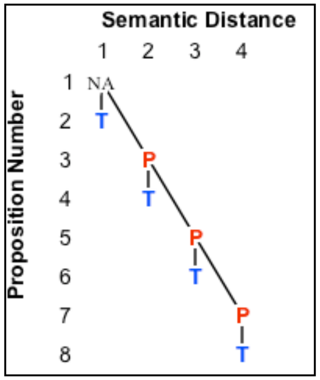

figure 5. Topic drift.

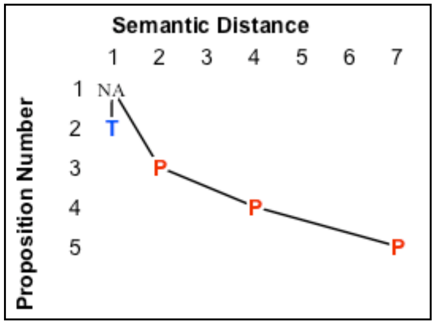

figure 6. Digression.

Making available a visual "morphology" of this sort could make DTA diagrams more readily interpretable. This, in turn, would extend their usefulness both as a research tool and as a potential user interface for text chat systems.

Conclusion

VisualDTA is a technique for visually representing interactional and pragmatic aspects of online discourse. As such, it goes beyond previous text chat visualization techniques that display purely structural information such as participant ID, message length, posting frequency, and time since last posting [1, 13]. It also shows the sequential relations between messages in synchronous CMC without requiring the use of threaded chat systems [10], making it usable with popular chat modes such as IM and IRC.

At present, DTA is limited in that it requires a human being to code inter-message relatedness and semantic distance. Despite promising work involving machine learning [12], automatic identification of semantic and pragmatic meaning remains a desideratum in computer-mediated discourse analysis, and in natural language processing more generally [7]. Future work should concentrate efforts in this area. Automating the semantic and pragmatic analysis in DTA would open the door to using VisualDTA to display the topical dynamics of textual conversations to users in real time.

Acknowledgments

We are grateful to Narayanan Iyer of Indiana University for collecting and making available to us the Instant Messaging data.

References

[1] Erickson, T. Smith, D. N., Kellogg, W. A., Laff, M. R., Richards, J. T., and Bradner, E. Socially translucent systems**:** Social proxies, persistent conversation, and the design of 'Babble.' In *Proceedings of CHI '99*, ACM Press (1999).

[2] Francis, G., and Hunston, S. Analyzing everyday conversation. In M. Coulthard (Ed.), *Advances in Spoken Discourse Analysis*, Routledge (1992), 1-34.

[3] Herring, S. C. Dynamic topic analysis of synchronous chat. In *New Research for New Media: Innovative Research Methodologies Symposium Working Papers and Readings*, University of Minnesota School of Journalism and Mass Communication (2003). http://ella.slis.indiana.edu/~herring/dta.html

[4] Herring, S. C., and Nix, C. G. *Is "serious chat" an oxymoron? Pedagogical vs. social uses of Internet Relay Chat.* Paper presented at the American Association of Applied Linguistics (1997).

[5] Hobbs, J. Topic drift. In B. Dorval (Ed.), *Conversational Organization and its Development,* Ablex (1990), 3-22.

[6] Kurtz, A. J., and Herring, S. C. *VisualDTA* (2004). http://ella.slis.indiana.edu/~ajkurtz/research/VisualDTA/

[7] Marcu, D. The rhetorical parsing of natural language texts. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics* (1997), 96-103.

[8] Schegloff, E. A., and Sacks, H. *Opening up closings*. Semiotica, 25 (1973), 289-327.

[9] Sinclair, J., and Coulthard, M. *Towards an Analysis of Discourse*, Oxford University Press (1975).

[10] Smith, M., Cadiz, J. J., and Burkhalter, B. Conversation trees and threaded chats. In *Proceedings of CSCW* (2000).

[11] Stromer-Galley, J., and Martinson, A. M. *Coherence in CMC: Comparing topics in chat* (2005). http://www. albany.edu/~jstromer/CoherenceinChatDRAFT.pdf

[12] Twitchell, D. P., and Nunamaker, J. F., Jr. Speech act profiling: A probabilistic method for analyzing persistent conversations and their participants. In *Proceedings of HICSS-37* (2004).

[13] Viégas, F. B., and Donath, J. S. Chat Circles. In *Proceedings of CHI '99*, ACM Press (1999), 9-16.