



# Understanding the dynamics of collaborative multi-party discourse

Andrew J Cowell<sup>1</sup>  
Michelle L Gregory<sup>1</sup>  
Joe Bruce<sup>1</sup>  
Jereme Haack<sup>1</sup>  
Doug Love<sup>1</sup>  
Stuart Rose<sup>1</sup>  
Adrienne H Andrew<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA, U.S.A.

Correspondence:  
Andrew J. Cowell, Pacific Northwest  
National Laboratory, P.O. Box 999, MS  
K7-28, Richland, WA 99352, U.S.A.  
Tel: +1 509 375 4548;  
Fax: +1 509 375 3641;  
E-mail: andrew@pnl.gov

## Abstract

In this paper, we discuss the efforts underway at the Pacific Northwest National Laboratory in understanding the dynamics of multi-party discourse across a number of communication modalities, such as email, instant messaging traffic and meeting data. Two prototype systems are discussed. The Conversation Analysis Tool (ChAT) is an experimental test-bed for the development of computational linguistic components and enables users to easily identify topics or persons of interest within multi-party conversations, including who talked to whom, when, the entities that were discussed, etc. The Retrospective Analysis of Communication Events (RACE) prototype, leveraging many of the ChAT components, is an application built specifically for knowledge workers and focuses on merging different types of communication data so that the underlying message can be discovered in an efficient, timely fashion. *Information Visualization (2006) 5, 250–259.*

doi:10.1057/palgrave.ivs.9500139

**Keywords:** Anthroposemiotics; discourse analysis; sentiment analysis; group dynamics; conversation topic segmentation

## Introduction

Communication is the heart of what makes us social creatures. Today, a myriad of technologies allow us to communicate in ways our forefathers could never have imagined. Computationally supported modalities such as email and instant messaging have had immeasurable effect on the way we work, play and interact with those in our lives. Being able to understand how individuals communicate, the methods they use, their personal preferences, etc. are all part of a field called ‘anthroposemiotics’.<sup>1</sup> This field looks to uncover the mystery behind how we communicate with ourselves (intrapersonal communication), with others (interpersonal communication), within groups (group dynamics) and across cultures (cross-cultural communication). While a great deal of literature exists in each of these fields,<sup>2–7</sup> there are few operational prototypes that allow for true hands-on investigation. Here, we discuss two projects underway within the Rich Interaction Environments group of the Pacific Northwest National Laboratory.

## The conversation analysis tool

The ability to extract and summarize content from data is a fundamental goal of computational linguistics. As such, a number of tools exist to automatically categorize, cluster and extract information from documents. However, these tools do not transfer well to data sources that are more conversational in nature, such as multi-party meetings, telephone conversations, email, chat rooms, etc. Given the plethora of these data sources, there is a need to be able to quickly and accurately extract and process pertinent information without having to cull them manually.

Much of the work on computational analysis of dialogue has focused on automatic topic segmentation of conversational data, and in particular, using features of the discourse to aid in segmentation.<sup>8–10</sup> Detailed discourse and conversational analytics have been the focus of much linguistic research and have been used by the computational community for creating models of dialogue to aid in natural language understanding and generation.<sup>11–13</sup> There has been a recent surge of computational discourse processing, especially with a focus towards automatic meeting summary tools. As such, there have been a number of efforts to collect meeting data, such as the Augmented Multi-party Interaction (AMI) Meeting Corpus,<sup>14</sup> the National Institute of Standards and Technology (NIST) Meeting Room Pilot Corpus,<sup>15</sup> the Interactive Systems Laboratories (ISI) Meeting Corpus<sup>16</sup> and the International Computer Science Institute (ICSI) Corpus.<sup>17</sup> A number of tools have also been developed to support the analysis of meeting corpora, such as NOMOS,<sup>18</sup> the NITE XML Toolkit<sup>11</sup> and CALOS.<sup>19</sup> The current work is most similar to tools designed to analyze multimodal conversational data such as NOMOS and the NITE XML Toolkit. Both of these tools were designed for annotating conversational data. NITE XML is a system in which transcripts of conversations are viewable and time aligned with their audio transcripts. It is especially useful for adding annotations to multi-modal data formats. NITE XML is not an analysis tool, however. Annotations are generally manually added. This is also true of NOMOS, which provides an environment for users to provide annotations to data, whether from audio files or text. Users can annotate topic breaks and run various other processing tools. These tools were designed mostly to integrate various sources of annotations of conversational data and so far have limited applications to meeting analysis. Much less focus has been on computational tools that aid in either analyzing conversations themselves or rendering conversational data so that it can be used with traditional data mining techniques that have been successful for document understanding.

In this paper, we present a ChAT that integrates several language-processing tools (topic segmentation, affect scoring, named entity extraction) that can be used to automatically annotate conversational data. The processing components have been specially adapted to deal with conversational data. ChAT is not intended as solely a meeting analysis tool, or meeting summarization tool, nor does it aid in annotating data. Instead, ChAT is an analysis tool in which the various language-processing components are fully automated. Its format makes it useful for analyzing any threaded dialogue, including instant messaging traffic, chat room data, phone transcripts and meeting transcripts: any kind of textual data where there is more than one participant.

ChAT includes a basic user interface that combines a variety of data sources onto one screen that enables users to progressively explore conversational data. For example, one can explore who was present in a given conversation, what they talked about, and the emotional content of the

data. The data can be viewed by time slice or in a semantic graph. The versatile language-processing components in ChAT were developed in modular, open designs so that they can be used independently or be integrated into other analytic tools (such as RACE, discussed later).

Data are brought in via an ingest engine. The central processing engine then normalizes the format (time stamp, speaker ID, utterance; one utterance per line). Processing components are called by the central processing engine that provides the input to each component and collects the output to send to the user interface.

The system was designed to handle multiple data types. Thus, with the exception of the ingest engine, the processing components are domain and source independent. For example, we did not want the topic segmentation to rely on features specific to a data set, such as acoustic information from transcripts. Also, all processing components have been built as independent plug-ins to the processing engine: the input of one does not rely on the output of the others. This lets users choose to include or exclude various processes to suit their needs or even exchange the components with new tools. We describe each processing component in the next section.

### Ingest engine

The ingest engine inputs multiple data sources and transforms them into a uniform structure that includes one utterance per line, including time stamp and participant information. So far, three data sources have been ingested. The ICSI meeting corpus<sup>17</sup> is a corpus of text transcripts of research meetings. Seventy-five meetings are in the corpus, each lasting 30–90 min, with 5–8 participants. A subset of these meeting transcripts was hand coded for topic segments.<sup>8</sup> We also used telephone transcripts from the August 14, 2003 power grid failure that caused a regional blackout.<sup>20</sup> These consist of files containing transcripts of multiple telephone conversations between multiple parties. Lastly, we employed a chat room data set that was built in-house by summer interns who were instructed to play a murder mystery game. Participants took on a character persona and content was based on a predefined scenario, but all interactions were unscripted beyond that.

### Topic segmentation

The output of the ingest process is a list of utterances that include a time (or sequence) stamp, a participant name and an utterance. Topic segmentation is then performed on the utterances to chunk them into topically cohesive units. Traditionally, algorithms for segmentation have relied on textual cues.<sup>21–24</sup> These techniques have proved useful in segmenting single-authored documents that are rich in content and have much topic continuity. Topic segmentation of conversational data is much more difficult due to often sparse content, intertwined topics and lack of topic continuity.

Topic segmentation algorithms generally rely on a lexical cohesion signal that requires smoothing to eliminate noise from word choice changes in adjoining statements that do not indicate topic shifts.<sup>23,25</sup> Many state-of-the-art techniques use a sliding window for smoothing.<sup>8,23,24</sup> We employ a windowless method (WLM) for calculating a suitable cohesion signal that does not rely on a sliding window to achieve the requisite smoothing for an effective segmentation. Instead, WLM employs a constrained minimal-spanning tree (MST) algorithm to find and join pairs of elements in a sequence. In most applications, the nearest-neighbor search used by an MST involves an exhaustive,  $O(N^2)$ , search throughout all pairs of elements. However, because WLM only requires information on the distance between adjoining elements in the sequence, the search space for finding the two closest adjoining elements is linear,  $O(N)$ , where  $N$  is the number of elements in the sequence. We can therefore take advantage of the hierarchical summary structure that an MST algorithm affords while not incurring the performance penalty.

Of particular interest for our research was the success of WLM on threaded dialogue. We evaluated WLM's performance on the ICSI meeting corpus<sup>17</sup> by comparing our segmentation results to the results obtained by implementing LCseg.<sup>8</sup> Using the 25 hand-segmented meetings, our algorithm achieved a significantly better segmentation for 20 out of 25 documents.

Topic segmentation of conversational data can be aided by employing features of the discourse or speech environment, such as acoustic cues, etc.<sup>8,10</sup> In this work, we have avoided using data-dependent features (the integration of acoustic cues for speech transcripts, for example) to aid in segmentation because we wanted our system to be as versatile as possible. This approach provides the best segmentation possible for a variety of data sources, regardless of data type.

### Named entity extraction

In addition to topics, ChAT also has integrated software to extract the named entities: Cicero Lite from the Language Computer Corporation (LCC) was used for entity detection (for a product description and evaluation, see<sup>26</sup>). Using a combination of semantic representations and statistical approaches, Cicero Lite isolates approximately 80 entity types. ChAT currently makes use of only a handful of these categories but can easily be modified to include more. Because named entity extraction relies on cross-utterance dependencies, the main processing engine sends all utterances from a conversation at once rather than an utterance at a time.

### Sentiment analysis

In addition to topic and entity extraction, conversations can also be analyzed by who participated in them, their relationship to one another and their attitude toward

topics they discuss. In an initial attempt to capture participant attitude, we have included a sentiment analysis, or affect, component. Sentiment analysis is conducted via a lexical approach. The lexicon we employed is the General Inquirer (GI) lexicon developed for content analyses of textual data.<sup>27</sup> It includes an extensive lexicon of over 11,000 hand-coded word stems, and 182 categories, but our implementation is limited to positive (POS) and negative (NEG) axes. In ChAT, every utterance is scored for the number of positive and negative words it contains. We make use of these data by keeping track of the affect of topics in general, as well as the general mood of the participants.

### Participant roles

Analyzing conversations consists of more than analyzing the topics within them. Inherent to the nature of conversational data are the participants. Using textual cues, one can gain insight into the relationships of participants to each other and the topics. In ChAT, we have integrated several simple metrics as indicators of social dynamics among the participants. Using simple speaker statistics, such as number of utterances, number of words, etc., we can gain insight into the level of engagement of participants in the conversation. Features we use include the number of utterances, proportion of questions vs statements and proportion of 'unsolicited' statements (ones not preceded by a question mark).

We also use the same lexical resources as we use for sentiment analysis for indications of personality type. We use the lexical categories of strong, weak, power cooperative and power conflict as indicators of participant roles in the conversational setting. Thus far, we have not conducted any formal evaluation on the sentiment analysis with these data, but our initial studies of our POS and NEG categories show a 73% agreement with hand-tagged positive and negative segments on a different data set.

These simple statistics outlined here are not intended to provide a comprehensive analysis of discourse structure. Instead, they provide indications of a participant's general role in a conversation. When used within the interface we provide, a user can easily see which participant contributes most in the whole conversation, or for an individual topic within the conversation. For example, in the meeting corpus data we have looked at, it is easy to identify the participants leading the meeting (they contribute to all topics and ask the most questions). We can also glean information about social status. One participant usually rarely speaks but dominates the meetings when a particular individual is not present. While the approach we have taken here, of relying on automatic counts, certainly does miss nuances of conversations and interactions, it has the advantage of being able to process massive amounts of data and analyze them in a manner sufficient for the typical task assigned to an analyst.

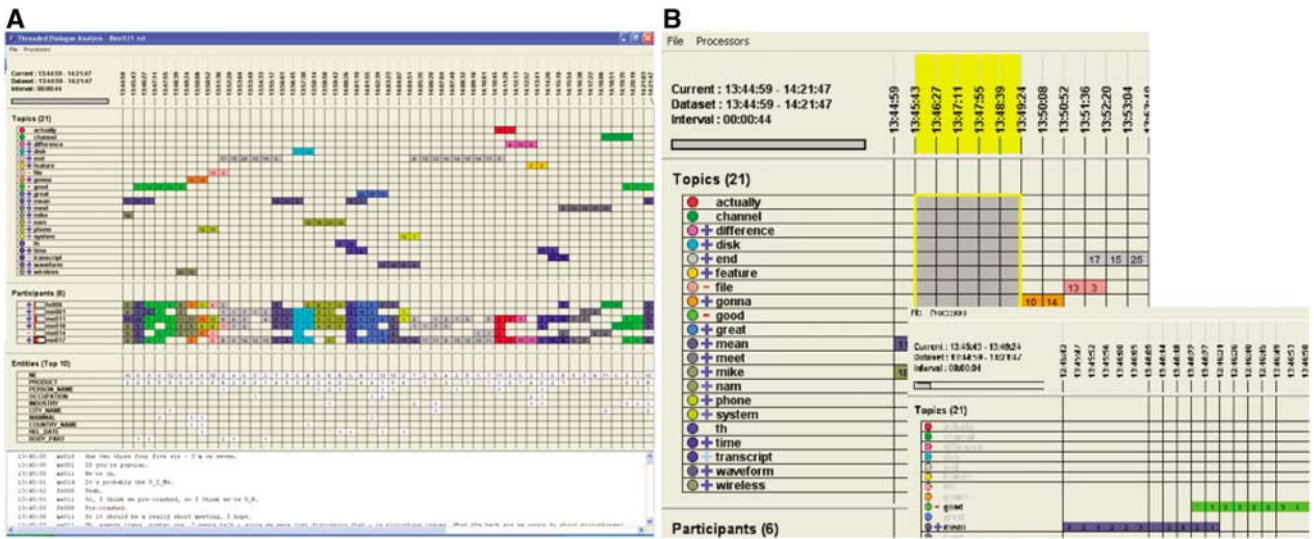


Figure 1 (A) The main ChAT and (B) time selection.

### User interface

As described earlier, the processing components are independent of the user interface. The user interface is designed solely as an analysis tool. It does not support any user annotations. ChAT is a stand-alone tool developed to provide 'broad brushed' information to analysts, with the capability to explore interesting relationships in more depth.

The components of the system are all linked through the X-axis, representing time, as seen in Figure 1A. Depending on the data set, positions along the time axis are based on either the time stamp or sequential position of the utterance. The default time range is the whole conversation or chat room session, but a narrower range can be selected by dragging in the interval panel at the top of the user interface. Note that all of the values for each of the components are recalculated based on the selected time interval. Figure 1B shows that a time selection results in a finer grained subset of the data, allowing one to drill down to specific topics of interest.

The number of utterances is indicated by the number inside the box corresponding to the time frame. The number is recalculated as different time frames are selected.

**Topics** The central organizing unit within the user interface is topics. The topic panel, illustrated in Figure 2A, consists of three parts: the color key, affect scores and topic labels. Once a data file is imported into the user interface, topic segmentation is performed on the data set according to the processes outlined earlier. Topic labels are assigned to each topic chunk. Currently, we use the most prevalent word tokens as the label, and the user can control the number of words per label. Each topic segment is assigned a color, which is indicated by the color key. The persistence of a color throughout the time axis indicates which topic is being discussed at any given time. This

allows a user to quickly see the distribution of topics of a meeting, for example. It also allows a user to quickly see the participants who discussed a given topic.

**Affect** Affect scores are computed for each topic by counting the number of POS and NEG affect words in each utterance that comprises a topic within the selected time interval. Affect is measured by the proportion of POS-NEG words in the selected time frame. If the proportion is greater than 0, the score is POS (represented by a +); if it is less than 0, it is NEG (-). The degree of sentiment is indicated by varying shades of color on the + or - symbol. Note that affect is computed for both topics and participants. An affect score on the topic panel indicates overall affect contained in the utterances present in a given time frame, whereas the affect score in the participant panel indicates overall affect in a given participant's utterances for that time frame.

**Participants** The participant panel (Figure 2B) consists of three parts: speaker labels, speaker contribution bar and affect score. The speaker label is displayed in alphabetical order and is grayed out if there are no utterances containing the topic in the selected time frame. The speaker contribution bar, displayed as a horizontal histogram, shows the speaker's proportion of utterances during the time frame. Non-question utterances are displayed in red while utterances containing questions are displayed in green. For example, in Figure 2B, we can see that speaker 'me011' did most of the talking (and was generally negative), but speaker 'me018' had a higher proportion of questions.

### Semantic graph

Data viewed in the main user interface can be sent to a semantic graph for further analysis. The graph allows a

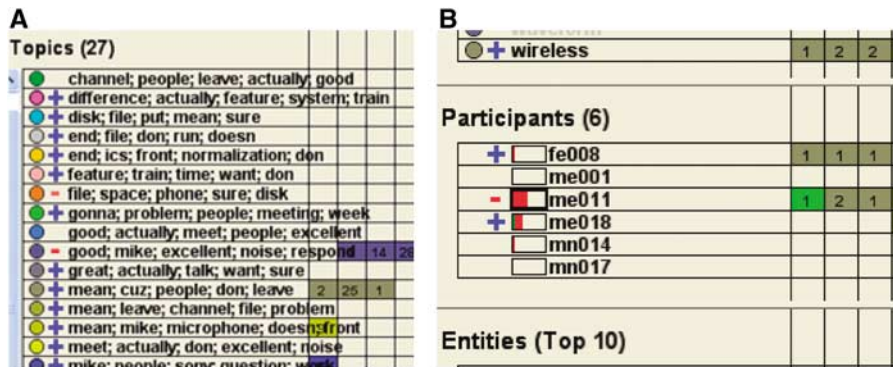


Figure 2 (A) Topic labels in the topic panel and (B) participant panel.

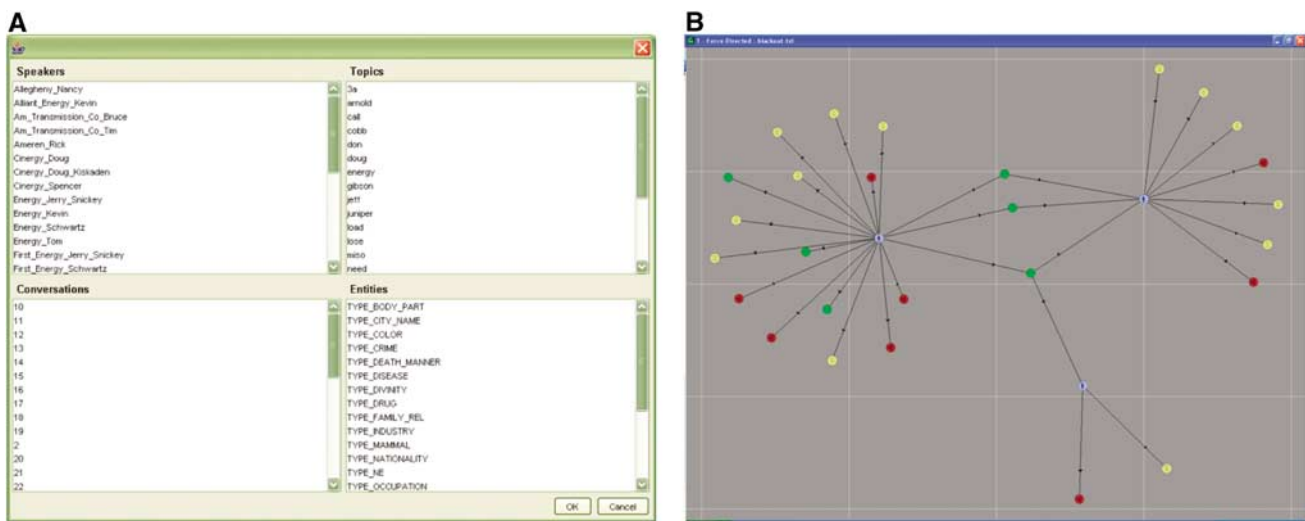


Figure 3 (A) Semantic graph node selection and (B) graph of the relationship between three participants.

user to choose to highlight the relationships associated with a topic, participant or individual named entity. The user selects objects of interest from a list (Figure 3A) and then the graph function organizes a graph according to the chosen object (3B), which extracts the information from the time-linked view and represents it in a more abstract view that denotes relationships via links and nodes.

The semantic graph can help highlight relationships that might be hard to view in the main user interface. For example, 3B represents a subset of the Blackout data in which three participants, indicated by blue, all talked about the same named entity, indicated by green, but never talked to each other, indicated by the red conversation nodes.

### Review

In this section, we presented ChAT, a system that helps analyze any kind of threaded dialogue. The processing components can be used independently or within the basic user interface. The components can also be run

independent of the user interface, in batch, resulting in an XML document containing the original transcripts and the metadata added by the processing components. This feature is not intended as an annotation tool, but rather to increase the functionality of the tool by allowing the data to be manipulated by traditional text mining techniques or to be viewed in other conversation analysis tools, such as RACE.

### RACE – retrospective analysis of communications events

A major focus of our work at the Pacific Northwest National Laboratory is providing valuable tools to knowledge workers. The area of communications analysis crosses many boundaries but perhaps nowhere is the application of this field more important than in the field of intelligence analysis.<sup>28,29</sup> Intelligence analysts must make sound judgments, coherently constructed from scattered heterogeneous fragments of information while being faced with significant time constraints. The information they use is rarely complete, often unreliable

and usually temporally and spatially diverse.<sup>30</sup> These dimensions need to be aligned and the information understood to enable the analyst to recognize sequences of inter-related events and hypothesize about future actions. The aim in RACE has been to aid the analyst by researching, designing and implementing, in conjunction with working analysts, a prototype for the investigation of collaborative, multi-party discourse. The focus is on reducing the complexity of analyzing communications data through a triage process — from a large corpus to a small handful of relevant conversations to finally a highly detailed view of one conversation, enhanced with socio-behavioral dimensions from ChAT. Below we present the design methodology and discuss the latest version of the prototype.

### Method

The design methodology used included a review of the literature, followed by in-depth focus group discussions with working analysts to determine requirements. Next, a participatory design process was used to gather more information from the user group, leading to a set of sketches. From these, an initial prototype was created. The test-bed is currently in its third phase of implementation and includes the integration of components (indicators of affect and social roles) developed for ChAT.

### Prior art

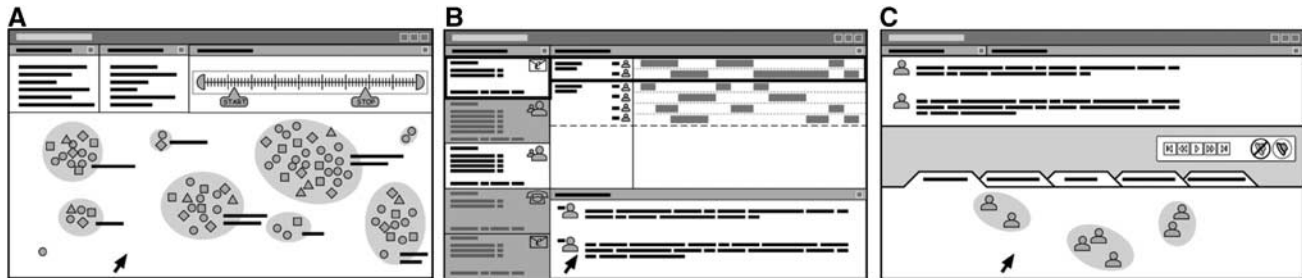
As both Internet communications and complex graphics capabilities have become more pervasive in modern computing, there has been much interest in visualizing conversations. Owing to the ease of data capture with computationally supported communications, such communication modalities as email, chat and forum/newsgroup threads appear to be the most researched. Several systems have represented vast, multi-threaded newsgroup or forum posts such as USENET. Loom can represent the activity patterns of individuals relative to one another, helping to characterize individuals' participation and roles.<sup>31,32</sup> In another view, linked posts are graphed to represent threads, characterizing the newsgroup as a whole. Discourse diagrams describe newsgroups with semantic graphs of related concepts and also graphs people's connectedness to one another in social networks.<sup>33</sup> Conversation thumbnails use an overview/detail display to contextualize a user's post in the group as a whole while it is being composed.<sup>34</sup> PeopleGarden represents each participant as a composite of his/her history of posting. Representing all participants in the same screen provides insight into the dynamics of the group as a whole across its recorded history, although there is no way to track connections between individuals or threading.<sup>35</sup>

In RACE, the goal is for an analyst to effectively look for both episodic and social information across the topics of a multitude of conversations. Through an

iterative filtering process, the analyst will examine individual conversations. Like the newsgroup visualizations above, the goals are (in addition to a general desire to understand the situation) to determine an individual's social role and dynamics of the group, but the concept of 'conversations' is more granular. Whereas the newsgroup visualizations may represent hundreds or even thousands of users and conversation threads, the detailed visualization in RACE's final screen represents a single discourse with as few as two people. Thus, the systems above deal with a higher level of abstraction and do not convey information on 'lurkers' who may read but not post, emotional qualities of contributions, or the temporal information present in synchronous communication. RACE has the additional goals of denoting presence, affect, and what Viegas and Donath call 'negotiation of conversational synchrony'.<sup>36</sup>

Research on chat room conversation has produced some interesting visualizations that start to deal with these concepts. Babble both facilitates and visualizes synchronous and asynchronous chat.<sup>37</sup> Users are represented as colored dots on a social proxy called a 'cookie'. The more interactions they have with the system, whether posting or only reading, the more central they become in the visualization. With inactivity, the dots move slowly back out to the periphery of the cookie, conveying information about presence and activity level. Chat circles is designed for synchronous chat and creates a strong sense of location by situating participants (represented as colored circles) in a large 2D space and only allowing them to see the text posted by others positioned nearby.<sup>36</sup> The circles expand to encompass posted text and shrink when ample time to read the utterance has passed. Even people who are idling or only listening are represented spatially so others can see them. People can position their circles to avoid the 'noise' of unrelated conversations (as one could do at a cocktail party) or signify whom they are addressing. Each post leaves a cumulative translucent trace, indicating how long the poster has been there and how active he/she has been. Thus, group dynamics such as a group conversation fragmenting into smaller ones, relative verbosity and relative position are available for interpretation.

While each of the systems above is designed for a particular modality, RACE integrates email, instant messaging, text messaging, in-person meetings, phone conversations and teleconferences in addition to chat and newsgroup participation. The goal is to get a more holistic sense of individuals throughout their discrete conversations and communication methods. As a post-hoc analysis tool, RACE aids the analyst by adding system interpretations of affect and social dynamics to the information represented in the prior art. It should be noted that this effort violates one of Erickson's six claims about social visualization: 'Portray actions, not interpretation... users understand the context better than the system ever will'.<sup>38</sup> We agree in theory, but the needs of our analysts differ from those of a contributor to the conversation. Content-driven interpretations of group dynamics, affect and social role complement full-text transcripts of the



**Figure 4** (A) Sketch of the corpus view and (B) sketch of the sequence view and (C) sketch of the detail view.

conversations, providing shortcuts to insight. Below we discuss further the requirements of our user group.

### Requirements elicitation

To ensure our research was applicable to our organization's missions and fulfilled the requirements and expectations of our user group, we enlisted the help of four analysts to determine specific requirements. These were to be our subject matter experts (SMEs). Through interactions with our SMEs, we determined that while it is important to understand a single conversation in time, it is just as, if not more, important to comprehend the stream of conversations that occurs over longer periods, related to the same topic. For example, it is important to be able to intercept, process and analyze a discussion between two individuals talking about making a homemade bomb, but it is even more important to place such a discussion within the context of the set of communications leading to an understanding of the overarching plot (note, this differentiates our work from meeting annotators, e.g., Abowd<sup>39</sup> and Brotherton and Abowd<sup>40</sup>, that aim to annotate a lecture or other event). Such review can provide additional information that could be invaluable to the analyst. Other requirements identified as part of these sessions included:

- The system should allow the analyst to get back to original source documents and review the provenance.
- The system should allow the analyst to annotate the communication events.
- Use of color for note taking and marking modalities should be considered.
- The system should allow the analyst to highlight conversation fragments (i.e., small parts of a larger conversation that are considered important).
- The system should provide basic translation mechanisms for foreign language support and some form of lexicon for terms that fall outside an analyst's field of expertise.
- The system should be able to import and export conversation fragments using common formats.
- The system should allow multiple analysts to work collaboratively within the same workspace.

- The system should allow the analyst to customize the environment to their preferences.

In addition to an informal list of requirements, a great deal of brainstorming was performed during this session. Following a participatory design process, system designers worked with SMEs to put together a work process and some initial sketches of the overall system that could be fed into the implementation stage.

The process (Figure 4A) was designed so that analysts could interact with the conversation corpus available to them (potentially produced as a result of a search), viewing the conversations as dots, clustered around major topics. This view could be filtered based on time period, participants involved and communications modality used.

On selecting a subset of conversations to review further, analysts move to a second screen (the sequence view, Figure 4B) where they can analyze the conversations in relation to when they occurred (the view is reminiscent of Microsoft Project's Gantt view).

While icons and text will continue to depict the modality the conversation used, the focus at this level is of fusing the conversations to build a sequenced stream of communications traffic so the underlying thread or purpose can be understood. Finally (Figure 4C), conversations of specific interest to analysts can be pursued in further detail in a third screen (detail view). Here, the full transcript is displayed and can be 'played' utterance by utterance in real time. As each utterance is reached, a text-to-speech engine speaks the words, while a number of visual representations indicate social constructs such as social roles and the dynamics between the individuals.

### Implemented prototype

Using a participatory design process, informed by the sketches and requirements of our analysts and the limitations of current research systems, we implemented a three-screen prototype analytical environment that allows a user to visualize a large corpus of communications events.

The environment can run on three screens simultaneously, be split across three panes (useful for performing

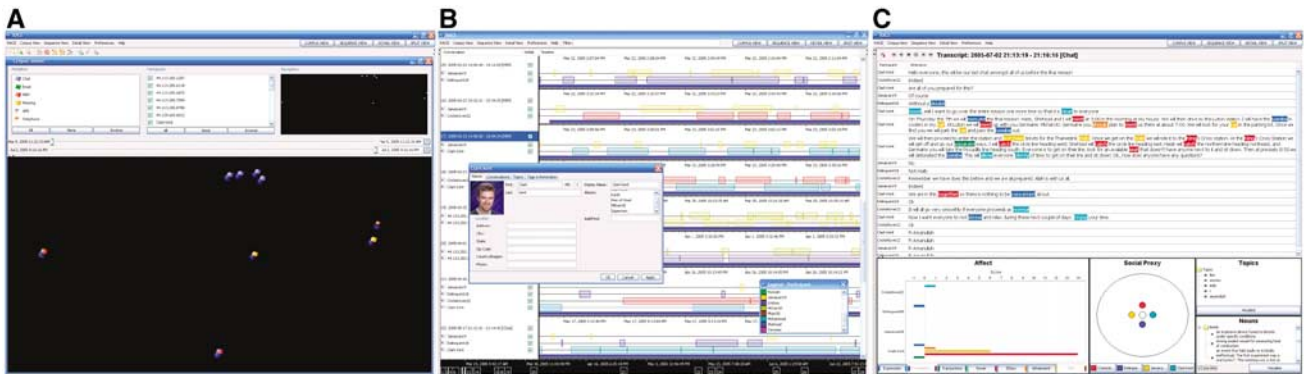


Figure 5 (A) The corpus view, (B) the sequence view and (C) the detail view.

analysis on large displays like wall-mounted plasma displays) or on a single screen with the use of a window manager seen in the top right of each view.

For the ‘corpus view’ (Figure 5A) we created a viewer for the IN-SPIRE<sup>41</sup> visualization suite to present the conversation corpus, clustered by topic. Zooming into individual items brings up metadata about that specific conversation. The different modalities may also be represented by different icons or colors, depending on the type of style sheet loaded. Filters currently available include the modality used, the participants involved and the time/date the conversation occurred (shortcuts to selecting all or none, or the current inverse are also available). Finally, a navigation window ensures that the user does not get lost when interacting with massive data that are topically diverse.

The ‘sequence view’ (Figure 5B) is where we envision analysts will spend the majority of their time. Here they will review, in detail, a small subset of conversations that they found of interest in the corpus space. For example, while exploring the visualization, analysts may find a group of discussions about a particular chemical substance. Knowing that this is relevant to a study they are performing, they simply drag a box around that subset and immediately those conversations are shown in the sequence view. Each conversation has an independent time line and can be zoomed out to show the entire conversation or zoomed in to see the individual utterances (these may also be accessed using tool-tips). The conversation titles on the left side of the screen can be expanded to show all the participants involved. Clicking a participant opens a dialog box containing known information about that individual (including any known aliases and other names he or she may use online). A global timeline at the bottom of the screen shows where each conversation falls in sequence.

Once an important conversation is uncovered through the triage process, it can be selected for deeper investigation in the details view (Figure 5C). This view can enable the analyst to see beyond the individual utterances. Using ChAT components, the details view lets the analyst gain insight into an individual’s opinion on the topics discussed. The transcript is color-coded to show

the seven dimensions of affect (expression, power, ethics, attainment, skill, accomplishment and transactions), while a graph representation allows the analyst to compare individuals’ affect against each other. To ingest the text in different ways, a ‘text-to-speech’ engine can be used to have the computer ‘speak’ the transcript. As it steps through the utterances, a group dynamics graphic (based on Erickson’s Social Proxy) shows how the individuals relate to each other, highlighting those involved in the conversation and those that are idle. This view also provides a hierarchical view of the topics discussed with the ability to trigger a multi-dimensional visualization that maps participants to topics.

### Evaluation and data sets

In addition to the prototype system, we developed an evaluation plan. The current data set being used to demonstrate the system was synthesized from news reports about the London bombings of July 7, 2005. The evaluation will use a new data set of telephone transcripts from the regional August 14, 2003 blackout<sup>20</sup> to ensure that any analysts who were involved in the development of the prototype will not benefit from any potential learning effects. These data are made up of several participants involved in many different conversations. These characteristics are exactly what RACE was designed for. Another data set is a transcript of a murder mystery enactment held on a chat room. While there was only a single space for characters to interact, there were many different threads of conversation going on at once. This data set will be useful for exploring the social dynamic part of RACE. We hope to show who the conversational ‘drivers’ were and explore what characteristics give someone away when hiding details they do not want other characters to discover.

### Further work

The ultimate goal of the RACE project is to assist analysts as they try to extract meaning from a myriad of sources. To this end, we started by talking with analysts themselves.

This is in recognition of the fact that no matter how powerful a tool might seem to its developers, it is useless unless the end users actually adopt it. By working with analysts every step of the way, we are keeping that goal in sight.

RACE's design as a test-bed enables other research to get in front of the analyst sooner. The quick insertion of the text affect work illustrates the capability to make functionality available to the user for evaluation. Showing analysts a concrete example of an idea allows them to get a better understanding of it and an easier way to elicit feedback for future work.

While this is an exciting first step, there are many avenues of crucial research still to be performed. In many fields, having access to all the communications events that occurred is rare. We need to research how to best enable the analyst to fill in these blanks. Potential approaches include hypothesized inference or the use of placeholders.

Currently, the prototype analytical environment only processes and displays textual transcripts of communication events. We decided to handle textual content first to ensure proof of principle before expending effort on the more challenging task of fusing video, audio, still images and text. Some effort has been expended on looking for suitable design metaphors that could aid an analyst in making sense of such diverse media (e.g., video production user interfaces such as Apple® Final Cut Pro®) but more research, design, and evaluation is required.

More effort needs to be expended on understanding how best to fuse different modalities of communication. Currently, a time-shifting approach is used to normalize an asynchronous email thread with similar-topic synchronous communications (e.g., telephone call, instant messaging session). This approach works but, to be successful, needs to be refined. At one level, the modality used is irrelevant – it is the essence of the event that is of primary concern. Being able to boil down the associated threads into one specific stream (e.g., multiple conversations across a number of modalities, all around the topic of plotting to explode a device at a particular location) is crucial to supporting the analytical tradecraft and allowing analysts to provide actionable intelligence to their superiors.

Conversations rarely keep to a single focused topic, and this can cause problems in the cluster visualization type approach used so far. Topic segmentation is a difficult research area and not one that we intend to pursue. A least three projects are currently underway at our institution that deal with this area. We plan to leverage their research results.

Finally, many elements of multi-party discourse exist outside linguistic boundaries. The words we use, how often we make an utterance, etc., all speak to who we are as individuals. While some of this is obvious and can be observed with just a cursory review of a transcript of the source material, other elements are discrete and hidden. For example, conversational statistics can be recorded and

used to determine an individual's level of engagement in a topic. Detection of familiarity (e.g., either by specific words not currently found in the present conversation or through the use of casual rather than formal speech) can indicate personal relationships between individuals in a dyad. Personality types can be inferred by markers indicative of leadership (e.g., number of interruptions performed/received, ability to change topic, use of power terms) or weaker, subversive roles (e.g., use of weak terms, submission of floor, deference to others). Analysts are rarely able to access such rich personality profiles of their subjects without performing an exhaustive analysis or calling in specialized help. While we are just beginning to integrate certain elements of social discourse, there are many other dimensions to be considered.

## References

- 1 Budd R, Ruben B. *Approaches to Human Communication*. Spartan Books: New York, 1972.
- 2 Bion WR. *Experiences in Groups: And Other Papers*. Basic Books: New York, NY, 1961.
- 3 Knapp ML, Vangelisti AL. *Interpersonal Communication and Human Relationships*. 5th edn. Allyn & Bacon: Boston, MA, 2004.
- 4 Lustig MW, Koester J. *Intercultural Competence: Interpersonal Communication Across Cultures*, 4th edn. Allyn & Bacon: Boston, MA, 2002.
- 5 Ruesch J, Bateson G. *Communication: The Social Matrix of Psychiatry*. W W Norton & Co: New York, 1968.
- 6 Gudykunst W, Ting-Toomey S, Chua E. *Culture and Interpersonal Communication*. Sage: Newbury Park, CA, 1988.
- 7 Tuckman B. Developmental sequence in small groups. *Psychological Bulletin* 1965; **63**: 384–399.
- 8 Galley M, McKeown K, Fosler-Lussier, E and Jing H. Discourse segmentation of multipart conversation. In: *41st Annual Meeting of the Association for Computational Linguistics* 2003 (Sapporo, Japan).
- 9 Hirschberg J, Nakatani C. A prosodic analysis of discourse segments in direction-giving monologues. In: *34th Annual Meeting for Association for Computational Linguistics* 1996 (Santa Cruz, California, USA).
- 10 Stolcke A, Shriberg E, Hakkani-Tur D, Tur G, Rivlin Z and Sonmez K. Combining words and speech prosody for automatic topic segmentation. In: *DARPA Broadcast News Workshop* 1999 (Herndon, Virginia, USA).
- 11 Carletta JC, Kilgour J. The NITE XML toolkit meets the ICSI meeting corups: import, annotation, and browsing. *Workshop on Machine Learning for Multimodal Interaction* 2005 (Martigny, Switzerland), Springer-Verlag; Berlin, 2005.
- 12 Core MG, Allen JF. Coding dialogs with the DAMSL annotation scheme. In: *AAAI Fall Symposium on Communicative Action in Humans and Machines* 1997 (Boston, MA).
- 13 Deemter V, Krahmer E, Theune M. Real Versus template-based natural language generation: a false opposition? *Computational Linguistics* 2005; **31**: 15–24.
- 14 McCowan I, Carletta J, Kraaij W, Ashby S, Bourban S and Flynn M. The AMI meeting corups. In: *5th International Conference on Methods and Techniques in Behavioral Research* 2005 (Wageningen, The Netherlands).
- 15 Garofolo JS, Laprun CD, Michel M, Stanford VM and Tabassi E. The NIST meeting room pilot corpus. In: *4th International Conference on Language Resources and Evaluation* 2004 (Lisbon, Portugal).
- 16 Burger S, MacLaren V, Yu H. The ISL meeting corpus: the impact of meeting type on speech style. In: *International Conference on Spoken Language Processing* 2002 (Denver, CO, USA).
- 17 Janin A, Baron D, Edwards J, Ellis D, Gelbart D, Morgan N, Peskin B, Pfau T, Shriberg E, Stolcke A and Wooters C. The ICSI meeting corpus. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 2003 (Hong Kong).

- 18 Niekrasz J, Gruenstein A. NOMOS: A semantic web software framework for annotation of multimodal corpora. In: *5th Conference on Language Resources and Evaluation 2006* (Genoa, Italy).
- 19 Niekrasz J et al. Ontology-based discourse understanding for a persistent meeting assistant. In: *AAAI Spring Symposium 2005* (Stanford, CA, USA).
- 20 Tauzin WJ. Blackout 2004: How Did it Happen and Why? *Hearings before the Committee on Energy and Commerce House of Representatives One Hundred Eighth Congress, First Session, September 3 and September 4, 2003*. US Government Printing Office: Washington.
- 21 Beeferman D, Berger A, Lafferty J. Statistical models for text segmentation. *Machine Learning* 1999; **34**: 177–210.
- 22 Choi F. Advances in domain independent linear text segmentation. In: *North American Chapter of the Association for Computational Linguistics 2000* (Seattle, WA, USA).
- 23 Hearst MA. TextTiling: segmenting text into multiparagraph subtopic passages. *Computational Linguistics* 1997; **23**: 33–64.
- 24 Miller NE, Wong P, Brewster M and Foote H. Topic islands – a wavelet-based text visualization system. *IEEE Visualization '98 1998* (Research Triangle Park, NCA, USA); 189–196.
- 25 Barzilay R, Elhadad M. Using lexical chains for text summarization. In: *ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization 1997* (Madrid, Spain).
- 26 Harabagiu S, Moldovan D, Clard C, Bowden M, Williams J and Bensley J. *Answer mining by com-bining extraction techniques with abductive reasoning*. In: *Twelfth Text Retrieval Conference 2003* (Gaithersburg, MD, USA).
- 27 Stone P. Thematic text analysis: new agendas for analyzing text content. In: Roberts C (Ed). *Text Analysis for the Social Sciences*. Lawrence Erlbaum Associates Inc., Mahwah, NJ. 1997;
- 28 Clark RM. *Intelligence Analysis – A Target-Centric Approach*. CQ Press: Washington, DC, 2004.
- 29 Lowenthal MM. *Intelligence – From Secrets to Policy*, 3rd edn. CQ Press: Washington, DC, 2006.
- 30 Krizan L. *Intelligence Essentials for Everyone*. Joint Military Intelligence College: Washington, DC, 1999.
- 31 Boyd D, Lee H-Y, Ramage D, Donath J. Developing legible visualizations for online social spaces. In: *Hawaii International Conference on System Sciences 2002* (Big Island, Hawaii, USA).
- 32 Donath J, Karahalios K, Viegas F. Visualizing conversations. In: *32nd Hawaii International Conference on System Sciences 1999* (Maui, Hawaii, USA).
- 33 Sack W. Discourse diagrams: interface design for very large scale conversations. In: *33rd Hawaii International Conference on System Sciences 2000* (Maui, Hawaii, USA).
- 34 Wattenberg MM, Millen DR. Conversation thumbnails for large-scale discussion. In: *Computer Human Interaction: Human Factors in Computing Systems 2003* (Fort Lauderdale, FL, USA).
- 35 Xiong R, Donath J. PeopleGarden: Creating Data Portraits for Users. In: *12th Annual ACM Symposium on User Interface Software and Technology 1999* (Asheville, NC, USA).
- 36 Viegas FB, Donath JS. Chat Circles. In: *Computer Human Interaction: Human Factors in Computing Systems 1999* (Pittsburgh, PA, USA).
- 37 Erickson T, Laff MR. The design of the 'babble' timeline: a social proxy for visualizing group activity over time. In: *Computer Human Interaction: Human Factors in Computing Systems 2001* (Seattle, WA, USA).
- 38 Erickson T. Designing visualizations of social activity: six claims. In: *Computer Human Interaction: Human Factors in Computing Systems 2003* (Fort Lauderdale, FL, USA).
- 39 Abowd GD. Classroom 2000: an experiment with the instrumentation of a living educational environment. *IBM Systems Journal, Special issue on Pervasive Computing* 1999; **38**: 508–530.
- 40 Brotherton JA, Abowd GD. Rooms take note: rooms takes notes. In: *AAAI '98 Spring Symposium 1998* (Boston, MA, USA).
- 41 Hetzler E, Turner A. Analysis experiences using Information Visualization. *IEEE Computer Graphics and Applications* 2004; **24**: 22–26.