



DIGITAL FORMATIONS

IT and New Architectures in the Global Realm

Edited by

Robert Latham and Saskia Sassen

Discourse Architecture and Very Large-scale Conversation

WARREN SACK

HISTORICALLY, NEW spaces for public discussion have been invented every few centuries (the agora, plaza, town square, town hall, café, newspaper). The introduction of electrical and electronic technologies in the twentieth century accelerated the rate of change in public spaces to a pace measured in decades (film, radio, television). Now with the increasing ubiquity of computer networks, new spaces for public discussion and exchange are invented, introduced, and updated on an almost continual basis (e-mail, newsgroups, Internet Relay Chat (IRC), weblogs, instant messaging, Napster, Gnutella).

This exponential increase in the rate of change has reached the escape velocity of the disciplines and professions normally accorded the responsibility to design, build and analyze public spaces. No longer is it only architects, civil engineers, and urban planners who design spaces for public discussion. Symptomatic of this transformation is a proliferation of new architectures of computers and networks that are not designed by traditional architects, for example, computer architectures, network architectures, and information architectures. Conversely, traditional architecture has become increasingly involved in efforts to extend its methodologies to cover computer networks by rendering them as “cyberspaces.”

The gaps between discourse, code, and architecture have now been bridged to the extent that it is crucial for us to understand issues such as the legal ramifications of network architectures on free speech.¹ Today public spaces for discussion include bits as well as bricks and boards. This convergence of language and architecture has frequently produced an assemblage that fails like the Tower of Babel. Discourse specialists (linguists, sociologists, legal scholars, political scientists) have not often enjoyed the reputation of great designers of spaces and architectures. On the other hand, artists, designers, engineers, and architects—renowned for their abilities to envision and execute the configuration and mixing of spaces

and materials—have often been typified as inept in the skills of writing and speaking. But we are now at a point in time when the future of the public space depends upon the ability to mix discourse and architecture in a new area of endeavor called *discourse architecture*.

Network architecture is the computer science of connecting machines to machines. Information architecture is primarily practiced by librarians and database and web designers to connect people to machines by making it easy for people to find information on networked machines. Discourse architecture is the practice of designing environments to connect people to people through networked computers. Or, more specifically, discourse architecture is the practice of designing networked environments to support conversation, discussion, and exchange among people.

Prior work in this area includes that of the original Discourse Architecture Laboratory, a research group at Apple Computer.² Closely related is a large variety of work in Computer-Supported Cooperative Work (CSCW), Computer-Human Interaction (CHI),³ and Computer-Mediated Communication (CMC). Most recently a number of research groups have emerged to focus on what has been called *social computing* and *social informatics*. Groups of this sort now exist at a number of industry research labs, universities and nonprofit organizations. Unlike many scholars who work in CSCW and CMC, researchers in the area of social computing have identified earlier work in architecture and urban design as useful and interesting for the design of networked spaces. Discourse architecture is an area of social computing in which environments for discussion are of primary importance.⁴

The practice of discourse architecture entails two kinds of work: one concerns the extension and use of methods from art and design; the second employs and further develops ideas from the humanities and social sciences. First, as a practice of design, discourse architecture concerns the design and implementation of new computer network technologies for discourse; that is, the means to shape the conversation that takes place within a given system. Just as physical architecture facilitates certain activities and inhibits others (compare, for instance, the exchanges sup-

² Founding members of this research group at Apple included Dave Curbow, Paul Dourish, Tom Erickson, Jed Harris, and Austin Henderson, with consulting help from Niklas Damiras, Sha Xin Wei, Brian Cantwell Smith, and Helga Wild. See <http://pliant.org> for more information about this group.

³ See especially Erickson, Herring, and Sack (2002) and Munro, Hook, and Benyon (1999).

⁴ The following definition of discourse architecture is a direct outgrowth of the writing I have done together with Susan Herring and Thomas Erickson. The following paragraphs should be compared with our co-authored work on the subject (Erickson, Herring, and Sack 2002). However, Herring and Erickson may not agree with the version that appears here.

¹ Cf. Lessig (1999).

ported by amphitheatres versus those supported by cafés), so do system architectures facilitate certain types of conversations. For example, media architectures like television broadcasting facilitate one-to-many exchanges but do not directly support a democratic, many-to-many exchange between people. In contrast, the Usenet newsgroup network protocol, for instance, does support many-to-many exchanges. Prior work exists in the fields of architecture, urban design, and the arts.

Second, the criteria for evaluating any given discourse architecture depends upon some means to critique the form, character, content, and extent of the supported discourse. Thus, discourse architecture is concerned with the structure of conversation itself; that is, with how the utterances of a conversation interrelate and build upon one another. Discourse architects are interested in analytical techniques for identifying conversational structure and explicating the forces that shape it. Relatively little research has been done to understand how network architectures influence existing patterns of discourse or facilitate new patterns. Furthermore, the work that has been done is spread across a wide array of humanities and social science disciplines such as linguistics, literature, theater, philosophy, anthropology, communications, computer science, information science, political science, psychology, rhetoric, and sociology and draws on diverse theories and methods. Consequently, the practice of discourse architecture entails the extension, synthesis, and production of new knowledge appropriate to disciplines of the social sciences, arts, and humanities.

This chapter is an introduction to discourse architecture. It is an introduction by example. First a new area of discourse is identified; an area that will be referred to as *very large-scale conversation* (VLSC). It is usually conducted on the Internet through the exchange of e-mail. VLSC facilitates many-to-many exchanges among citizens across international borders. I argue that VLSC poses a fundamental challenge to existing social science methodologies because it constitutes a different scale of conversational interaction, a scale that has not been previously addressed by social science. I propose a computationally enabled means to understand and theorize VLSC and illustrate this proposal with a prototype piece of software, the Conversation Map. Finally, I argue that the Conversation Map is not just a tool but also a *technology of the self*, a means of self-reflection.

Very Large-scale Conversation

On the Internet there are now very large-scale conversations in which hundreds, even thousands, of people exchange messages across interna-

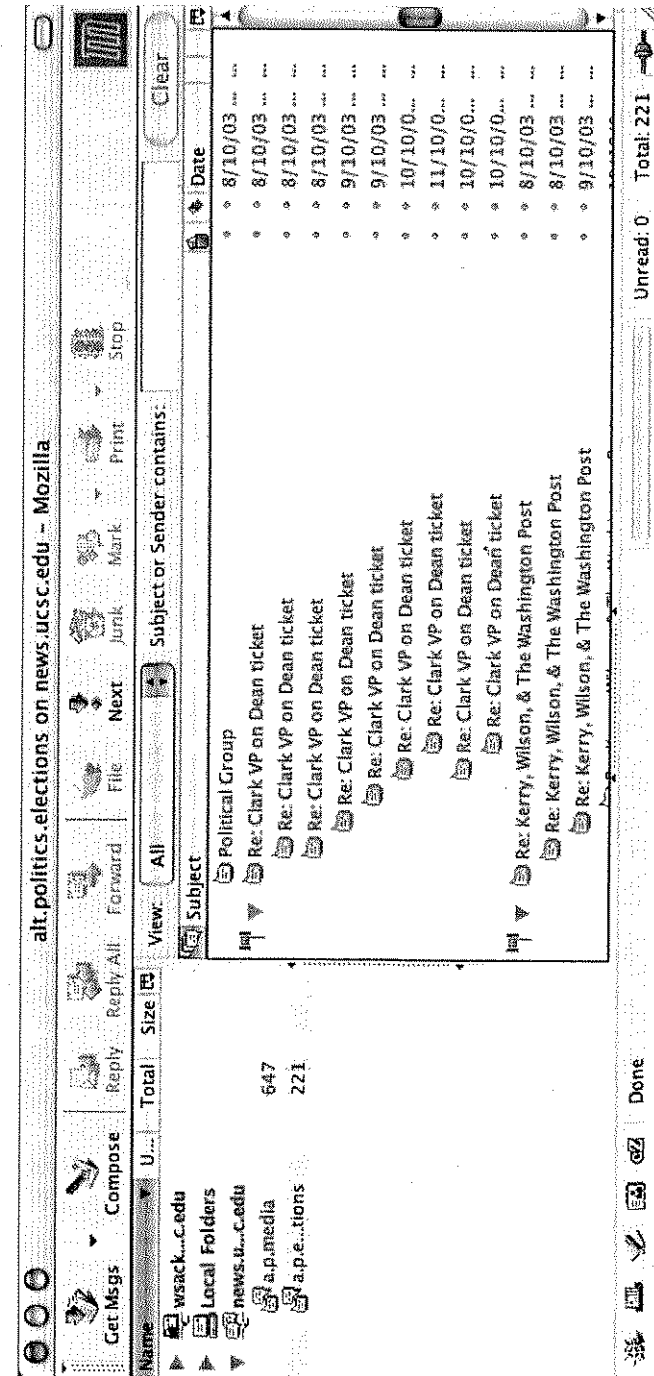


Figure 1. Mozilla News—a typical, contemporary view of VLSC

tional borders in daily, many-to-many communications. VLSC is an emergent communication medium that engenders new social and linguistic connections among people. It poses fundamental challenges to the analytical tools and descriptive methodologies of social science previously developed to understand conversations of a much smaller scale.

VLSC is both a well-known phenomenon and, simultaneously, something as yet largely unexamined by designers and social scientists. On the one hand, VLSC is well known in the form of busy Usenet newsgroups and large, electronic mailing lists and weblogs.⁵ For participants and observers alike, VLSC manifests itself as huge lists of messages in a conventional e-mail reader like RN, Eudora, or Netscape Messenger.

On the other hand, VLSC is largely unexamined. What does it mean to have a conversation that involves hundreds or thousands of people? Existing theories of conversation and discourse do not cover this scale of conversation. Moreover, very little design work for VLSC has been done. For example, why is VLSC usually represented as a long list of e-mail messages? Isn't something better possible? In fact, with a better theory of VLSC, better software for navigating VLSCs can be designed.

Detailed, micro analyses of face-to-face conversation usually involve a very different kind of work and produce a very different type of research result—that is, a very different type of knowledge—than do macroscale analyses of discourses involving thousands or millions of people. This micro/macro divide is a recurrent one in many of the social sciences and has been widely discussed in, for example, economics and sociology. Bridging this divide for the analysis of VLSCs is necessary because, on the one hand, the phenomenon under examination is macroscale by definition, but, on the other hand, one of the most ethically important motivations for analyzing VLSCs is to give participants a means to find their way and locate their position in a VLSC. Consequently, standard social scientific methods of dealing with macroscale phenomena by working with norms and averages are unworkable because they risk effacing the contributions of particular individuals.

I will argue that a bridge can be found between micro- and macroscale analyses of online conversations. This bridge is the lexicon or what might be called the “thesaurus” of a group conversation. On the micro scale, contributions to a conversation are judged to be coherent and cohesive partially according to whether or not they are taken to be “on topic” by the participants. Knowledge of deviation or convergence with a given topic is based on knowledge of a lexicon; that is, according to the rela-

⁵ Readers unfamiliar with these forms of online exchange might be interested in exploring <http://www.google.com/grphp>, an index of tens of thousands of Usenet newsgroups; <http://groups.yahoo.com/>, a selection of tens of thousands of web-based discussion groups; and/or <http://blogdex.net/>, an index of thousands of weblogs.

tionships between and the definitions of words. But, over the course of the lifetime of a group, new (e.g., slang) words are coined, some words gain new meanings, and others lose their currency, their connotations, or the controversy that surrounds them. Thus, conversation both depends upon and changes the lexicon or “thesaurus” of a group.

This conceptualization of VLSC—as the substrate and catalyst of community—is concordant with a large amount of work in sociolinguistics and the sociology of knowledge. Roughly speaking, what characterizes many of these sociolinguistic and sociological approaches to conversation and discourse is this: through the production and reproduction of a way of speaking and/or writing about certain pivotal subjects, a group is formed and distinguishes itself from other groups. Thus, chemists in the eighteenth century distinguished themselves from alchemists by developing a new discourse that we now recognize as the science of chemistry. Rather than talking about water as an essential element, chemists talk of the combination of hydrogen and oxygen. So, a new way of speaking and writing simultaneously produces a new group (e.g., chemists) and unravels or divides itself from a preexisting group (e.g., alchemists).

A way of speaking and writing (re)produces limits and possibilities for the way a subject can be spoken and/or written about and, simultaneously, (re)produces a social structure (e.g., a group or community). This way of thinking about the process and product of verbal interaction is well known in, for instance, conversation analysis.⁶ This way of describing the product or production of written and conversational forms has been termed “a discourse” by various European “continental” theorists:

[C]ontinental discourse theorists such as Foucault, Lyotard, Donzelot, Pêcheux, and De Certeau tend to use the term “discourse” to refer to relatively well-bounded areas of social knowledge. So, at any given historical conjuncture, it is only possible to write, speak, or think about a given social object (madness, for example) in specific ways and not in others. “A discourse” would then be whatever constrains—but also enables—writing, speaking, and thinking within such specific historical limits. Thus while a discourse can be thought of as linguistic in one sense, it also has to be treated in terms of the conditions of possibility of knowing a specific social object. (McHoul 1994: 944–45)

⁶ Harvey Sacks, the inventor of conversation analysis, provides several examples of the coconstruction of a lexicon and a social group: “Now if we can take it that to some extent ‘hotrodders’ is a category that is by and large employed by kids to characterize themselves, and whose use, to some considerable extent, they enforce, and whose properties they enforce, and obviously it’s, at least to some extent, a category that rebellious persons can use, then at least one of the initial questions we might ask is: Why should it be the case that at least some people who go about doing kinds of rebellion, do it by formulating themselves as a particular type? That is, why do they set up a type? Why don’t they try to make themselves observable as ‘individuals,’ so to speak?” (1992: 1:172).

From this continental perspective it is therefore possible to talk about, for instance, “the discourse of chemistry.” This usage of the term “discourse” (i.e., the use of the term discourse preceded by a definite or indefinite article like “the” or “a”) is sometimes at odds with or appears more or less incomprehensible to practitioners of other sorts of Anglo-American forms of discourse analysis.⁷

In her book that compares and contrasts six different Anglo-American approaches to discourse analysis (speech act theory, pragmatics, ethnomethodology, interactional sociolinguistics, ethnography of communication, and variation theory), Deborah Schiffrin (1994: 339) states: “Discourse has often been viewed in two different ways: a structure, i.e., a unit of language that is larger than the sentence; and the realization of functions, i.e. as the use of language for social, expressive, and referential purposes.” In other words, from an Anglo-American perspective, “discourse” is a name for a sequence of sentences (a structure) or a certain kind of language use (a function). But from a European, continental perspective, “discourse” is the result of either language use or the background conditions or context for a given sequence of sentences. Borrowing the trope of “figure/ground” from art history, one might say the difference between scholarly approaches to discourse analysis arises from the use of the term discourse to describe *figure* versus its use to describe *ground*. Or, alternatively, the conflict involves the use of discourse as a name for *text* versus its use as a name for *context*.

Rather than sort out this knotty conflation and conflict of terminology, I will try to find a way around it. From a continental perspective, one might talk about how a VLSC produces or reproduces a given or new discourse. From an Anglo-American perspective, one might say that a VLSC is a discourse. Instead, I will simply state that a VLSC produces, reproduces, and relies on a set of social and semantic relationships. In the language of mathematics, one might say that there exists a mutually recursive relationship between a VLSC and a set of social and semantic networks. Or, one might say, the coherence of a VLSC depends upon social and semantic background knowledge, but this background knowledge is also, at least partly, a product of the VLSC.

Three Dimensions of Conversational Commonsense

For conversations of a smaller scale (i.e., smaller than VLSC), it is possible to see when the background knowledge of a conversation is being

⁷ Some of the practical implications of these incommensurable differences between Anglo-American and European approaches to discourse are described in Pennycock (1994).

abused or flouted. Commonsense, conversational background knowledge can be described in a variety of ways: as a set of common associations and common terms, as a set of social and semantic networks, or—as will be elaborated below—as a set of metafunctions named *interpersonal*, *textual*, and *ideational* by Michael Halliday (1994: 179).

Divergences or differences of routine, conversational background knowledge can produce misunderstandings and conflict, but they can also produce comedy. Consider the following one-liner from comedian Stephen Wright: “I was driving down the highway one morning and I saw a billboard advertising a restaurant that said ‘Breakfast any time’ so I stopped and ordered French toast in the Renaissance.” The social coherence of a group underwrites conversation and depends upon a number of things. Semantics is just one of these things, but Wright’s joke illustrates how the production of common terms—a shared semantics—is important to conversation.

If the terms of conversation are followed, but the conventional turn-taking “rules” are not, another sort of nonsense is produced. Lewis Carroll (1960: 97) illustrates the “rules” of riddles when he has the characters of *Wonderland* violate them:

“Have you guessed the riddle yet?” the Hatter said, turning to Alice again.

“No, I give it up,” Alice replied: “what’s the answer?”

“I haven’t the slightest idea,” said the Hatter.

“Nor I,” said the March Hare.

Alice sighed wearily. “I think you might do something better with the time,” she said,

“than wasting it in asking riddles that have no answers.”

The common terms and rules of conversation are tightly coupled in the production of the cohesion of a conversation. When the cohesion is deliberately undone, the conversation is unhinged, as this snippet from Eugene Ionesco’s (1958) absurdist play *The Bald Soprano* illustrates. Suddenly, in this dialogue sequence, all of the people being discussed are named Bobby Watson:

MRS. SMITH: But who would take care of the children? You know very well that they have a boy and a girl. What are their names?

MR. SMITH: Bobby and Bobby like their parents. Bobby Watson’s uncle, old Bobby Watson, is a rich man and very fond of the boy. He might very well pay for Bobby’s education.

MRS. SMITH: That would be proper. And Bobby Watson’s aunt, old Bobby Watson, might very well, in her turn, pay for the education of Bobby Watson, Bobby Watson’s daughter. That way Bobby, Bobby Watson’s mother, could remarry. Has she anyone in mind?

MR. SMITH: Yes, a cousin of Bobby Watson's.

MRS. SMITH: Who? Bobby Watson?

MR. SMITH: Which Bobby Watson do you mean?

MRS. SMITH: Why, Bobby Watson, the son of old Bobby Watson, the late Bobby Watson's other uncle.

Obviously, writers and comics know and bend the common terms and rules of conversation in order to produce these sorts of effects. Using insights of this sort, scholars like Roman Jakobson (1985) have been able to explain the linguistic workings of avant-garde artistic literature, but the wittiness of more common performances also often depends upon an explicit understanding of how conversation engenders social cohesion and/or how the norms can be manipulated to reveal or break the underpinnings of social cohesion.⁸ It is equally as obvious that anyone who finds

⁸ In linguistics there exists a principle called Ziff's Law (Ziff 1960). Ziff's Law is the observation that any arbitrary string can be interpreted as a proper name. This is often mentioned as a serious difficulty for the construction of computer programs to parse natural language texts. However, it is also the main observation underlying Bud Abbott and Lou Costello's famous "Who's on first?" skit first performed on the *Kate Smith Radio Hour* in 1936:

COSTELLO: Look Abbott, if you're the coach, you must know all the players.

ABBOTT: I certainly do.

COSTELLO: Well you know I've not met the guys. So you'll have to tell me their names, and then I'll know who's playing on the team.

ABBOTT: Oh, I'll tell you their names, but you know it seems to me they give these ball players now-a-days very peculiar names.

COSTELLO: You mean funny names?

ABBOTT: Strange names, pet names . . . like Dizzy Dean . . .

COSTELLO: His brother Daffy

ABBOTT: Daffy Dean . . .

COSTELLO: And their French cousin.

ABBOTT: French?

COSTELLO: Goofé

ABBOTT: Goofé Dean. Well, let's see, we have on the bags, Who's on first, What's on second, I Don't Know is on third . . .

COSTELLO: That's what I want to find out.

ABBOTT: I say Who's on first, What's on second, I Don't Know's on third.

COSTELLO: Are you the manager?

ABBOTT: Yes.

COSTELLO: You gonna be the coach too?

ABBOTT: Yes.

COSTELLO: And you don't know the fellows' names.

ABBOTT: Well I should.

COSTELLO: Well then who's on first?

ABBOTT: Yes.

COSTELLO: I mean the fellow's name.

ABBOTT: Who.

these manipulations funny or absurd has a set of well-developed intuitions about the rules and terms of conversation: the *commonsense knowledge of conversation*.

Each of the comedic examples above illustrates a different metafunction of language. According to Michael Halliday (1994: 179), language has at least three meta-functions: (1) *ideational*: language can represent ideas; (2) *interpersonal*: language functions as a medium of exchange between people; and, (3) *textual*: language functions to organize, structure, and hold itself together; this function allows the various devices of cohesion, including citation, ellipsis, and anaphoric reference, to be used. The Steven Wright joke shows what can happen when the ideational metafunction breaks down. The selection from *Alice in Wonderland* illustrates the breakdown of the interpersonal metafunction. And in Ionesco's dialogue, the textual metafunction is thwarted by a breakdown of lexical cohesion. The point of these examples is simply to give examples of what might be considered the three different dimensions of commonsense knowledge about conversations that must be in place for a conversation—and so, transitively, a group of interlocutors—to hold together.

When one or all of these sorts of conversational background knowledge fall apart, the result can be funny.⁹ But by citing only the absurd and the comedic, it is difficult to picture what can be lost if the terms or rules of conversation are questioned or broken. While these questions and breaks can be funny, they can also arouse anger or mistrust.

Harold Garfinkel (1967: 43–44) asked his students to document this, the breakdown of common terms assumed in conversation; that is, to document the breakdown of the *ideational* metafunction. In the course of everyday conversation, Garfinkel's students questioned the assumed, common terms. The results were graphic. In the following accounts, Garfinkel's students play the role of the so-called experimenter (E).

The subject was telling the experimenter, a member of the subject's car pool, about having had a flat tire while going to work the previous day.
(S) I had a flat tire.

COSTELLO: The guy on first.

ABBOTT: Who.

COSTELLO: The first baseman.

ABBOTT: Who.

COSTELLO: The guy playing . . .

ABBOTT: Who is on first!

COSTELLO: I'm asking you who's on first.

ABBOTT: That's the man's name.

⁹ The violation of these sorts of commonsense knowledge can be seen as funny, as can the violation of a large variety of everyday expectations. See Freud (1960).

(E) What do you mean, you had a flat tire?

She appeared momentarily stunned. Then she answered in a hostile way: "What do you mean, 'What do you mean'? A flat tire is a flat tire. That is what I meant. Nothing special. What a crazy question!"

...

"On Friday night my husband and I were watching television. My husband remarked that he was tired. I asked, 'How are you tired? Physically, mentally, or just bored?'"

(S) I don't know, I guess physically, mainly.

(E) You mean that your muscles ache or your bones?

(S) I guess so. Don't be so technical.

(After more watching)

(S) All these old movies have the same kind of old iron bedstead in them.

(E) What do you mean? Do you mean all old movies, or some of them, or just ones you have seen?

(S) What's the matter with you? You know what I mean.

(E) I wish you would be more specific.

(S) You know what I mean! Drop dead!

...

The victim waved his hand cheerily.

(S) How are you?

(E) How am I in regard to what? My health, my finances, my school work, my peace of mind, my . . . ?

(S) (*Red in the face and suddenly out of control.*) Look! I was just trying to be polite. Frankly, I don't give a damn how you are.

These examples make the risks clear. By questioning the common terms of conversation, the students threaten the social contracts, or at least the smooth functioning, of various small groups of people: the car pool, the marriage, the friendship.

Questioning the common terms—the ideational metafunction of language—has risks. Analogously, there are risks to questioning the textual and interpersonal metafunctions. Using an ethnographic methodology, John Gumperz and his colleagues have documented how the textual and interpersonal metafunctions of language can break down in cross-cultural conversational situations. Consider the following utterances spoken by a Malaysian-born Indian immigrant in a London Adult Education class discussion about mortgages: "Mortgages. If you are to buy a house. Who can get and who cannot get. What assumptions we made, what? If you work. If you don't work, can you get a mortgage?" Gumperz et al. comment on this example: "the difficulties here are in following the connections that are being made, and consequently in understanding the intention of the final questions. [The] example starts with a string of noun phrases that appear to announce the intended topics. Is the final question

intended to elicit a review of the assumptions made at another time, or is it the commencement of discussion of the topic of 'who can get and who cannot get' a mortgage?"

Of course, the "difficulties" that Gumperz et al. mention are their difficulties, not the difficulties of the speaker or her audience, who are also, largely, English-speaking Indian immigrants. Gumperz et al. show how the structures and resources of grammar, prosody, and intonation of Hindi, Urdu, Punjabi, Gujarati, and Marathi are employed to join together multiple sentences when speakers of these North Indian languages speak English. Thus, the difficulties in resolving the cohesion are mostly difficulties for Anglo-American English speakers, not English-speaking Indians and Pakistanis. Consequently, even in a situation where the language being spoken is English and everyone in the situation is perfectly fluent in English, cross-cultural ties cannot, at times, be created because the textual and interpersonal metafunctions are produced very differently by members of different cultural groups.

This can have grave repercussions in legal, medical, and employment situations. In such situations bilinguals are sometimes thought to not be telling the truth because their testimony seems to be self-contradictory when interpreted by monolinguals; or the bilingual does not receive the medical care they need because the doctor does not understand them; or the bilingual does not get the job because the monolingual thinks the bilingual is hard to understand. In other words, in such situations—unlike the example situations of Garfinkel—the social fabric of a group is not ripped; rather, the group or social relationship is never threaded together or is clipped off right from the start.

Obviously the ideational, interpersonal, and textual relations established through inter- and intracultural conversational interactions do not remain static. Some groups become closer knit over time. Others fall apart. Intercultural, multilingual interactions can produce creoles and new forms of intelligibility; or, unfortunately, such situations can deteriorate through repeated miscommunication, and so cross-cultural conversation can become more and more difficult. To understand these shifts it is necessary to understand how a series of conversational interactions add up and thereby influence the performance of the metafunctions of language. For instance, how can good first impressions make interactions thereafter easy? How can a set of misunderstandings lead to diminished rapport between people who have gotten along for years?

The Micro-Macro Divide

It is quite easy to roughly characterize the difficulties of visualizing VLSC as a substrate and catalyst for community. It is a "chicken and egg" prob-

lem. The communities of VLSC—and thus, also, the conversational common sense of the community—do not preexist the VLSC except in some very vague manner. The texture and ideas of online communities come through collective actions and individual interactions, but it is difficult to see how a multiplicity of such (inter)actions might add up to, for instance, a coherent conversation, or what in continental theories might be called a discourse. The difficulty is what social theorists often refer to as the *micro-macro problem*: how can a large number of individual interactions add up to a larger social or political force and, vice versa, how does a larger social force act on small-scale, even intimate interactions?

Social theory has been in general terms concerned with different levels of analysis. In economic theory we are familiar with the idea of micro and macro economics to describe these different levels. Micro economics is concerned with the economic activity of individual economic units such as the household. Macro economics considers the behavior of the economy as a whole. Political science and sociology also work with such a distinction. In commonsense terms the micro level is the level of everyday interaction typically involving face-to-face negotiation between individuals. By contrast the macro level refers to the global structure of societies, and the analysis of major institutions such as the interface between the economy and politics; it also deals with large-scale collective action such as global social movements. The majority of social theorists recognize implicitly some form of this distinction, and various social theories have attempted to explicate the relationship between the micro and macro levels. (Turner 1996: 222).

Just as there is a micro-macro divide in economics and sociology research, there is also a micro-macro divide in discourse and conversation analysis work. The great majority of work done on conversation, by linguists and sociologists, consists of micro analyses of interactions between a small group of people. For example, work in conversation analysis often examines interactions between two or three people (e.g., Sacks 1992). Larger-scale work includes analyses of individual classrooms or small group interactions involving ten, twenty, or thirty people (e.g., Sinclair and Coulthard 1975). But large-scale work in examining interactions among hundreds or thousands of people, for instance, in online news-groups or interchanges in scientific literatures, usually effaces so many of the rich language details that microanalyses take particular care with that these large-scale investigations are a completely different species of work. These studies are, in other words, macro analyses, and it is difficult to see whether or how they complement the work of micro analyses. Historically, the most expedient thing to do has been to choose either a micro- or a macroanalysis methodology and then ignore the results of the other. However, this is not an option for VLSC because it is large scale, thus

macro, in size, but its rich details are what makes it a conversation rather than just, for example, an “information superhighway.”

For instance, recent work by Steve Whittaker, Loren Terveen, Will Hill, and Lynn Cherny (1998) on the dynamics of massive interaction analyzes the headers (the *to:*, *from:*, *references:*, etc. forms) of several million e-mail messages to investigate online conversational dynamics, but they do this analysis by completely ignoring the contents of the messages. Arguably, this sort of methodology—like a lot of work in sociology on social networks (e.g., Wasserman and Galaskiewicz 1994) and cocitation analysis (e.g., Garfield 1979)—is an exploration of some of the interpersonal dimensions of the medium of VLSC, but it leaves untouched the textual and ideational relations established or broken by VLSCs. Since the production and reproduction of social groups through VLSC is a function of at least all three of these aspects of language (the interpersonal, the textual, and the ideational), a strictly social network-based examination (who is responding to whom) is not sufficient as a complement to detailed microanalysis work.

Conversely, much other large-scale work has been done on text corpora that reveals recurrent patterns of ideational and textual relations but ignores how a series of texts can produce or reinforce a social network, a set of interpersonal relationships. For example, corpus-based, computational linguistics work has developed technologies for automatically compiling rough-draft thesauri given a large archive of texts;¹⁰ or, given an archive of tagged and bracketed texts, machines have been developed to automatically generate a grammar and a parser;¹¹ or, given a set of texts that mention many of the same people or places, some newly developed machines can now automatically hyperlink the texts so that entities in one text are automatically connected to mentions of the same entities in other texts.¹² Many of these same techniques have been taken up by sociologists of science working in the area of actor-network theory (e.g., Law and Hassard 1999). For example, Geneviève Teil and Bruno Latour (1995) describe a machine that uses measurements of conditional probability and mutual information to automatically compile a rough-draft thesaurus from a corpus of scientific abstracts.¹³

The difficulty with visualizing the conditions and productions of VLSC is therefore the following. Even though it would be ideal to simply “scale-

¹⁰ For contemporary work, see Grefenstette (1994); Hearst (1998); Harabagiu and Moldovan (1999). For the history of this field, see, for example, Soergel (1974).

¹¹ See, for example, Magerman (1994).

¹² See Bagga, Baldwin and Shelton (1999); Green (1997).

¹³ Teil's and Latour's work is one of the latest outgrowths of a long line of such computerized text analysis work (on co-word analysis) conducted at the Centre de Sociologie de l'Innovation, Ecole des Mines de Paris. See also, for example Callon, Law and Rip (1986).

up” the methodologies of the micro analysis of conversations and discourse, such methodologies can no more be scaled-up than the rich insights into bird flight gathered by a keen-eyed ornithologist can be scaled-up to analyze the dynamics of jet airplane travel. This is because these microscale analyses require too much of the analyst. Often a micro analysis of a conversation demands that the analyst identify the intentions of the participants; this is nontrivial if not impossible to do for a discussion that involves thousands of people.

Moreover, previous attempts to create theoretical tools for the examination of large text corpora have often neglected one or another metafunction of language (e.g., the interpersonal, the textual, the ideational) that is clearly important for conversational interaction. These large-scale theories of language are not adequate as large-scale theories of conversation because they leave out too many details. Large-scale work tends to fall into either (a) a social network type of work, which usually leaves out a lot about the contents of the text or talk exchanged; or (b) a corpus-based linguistics style of work that tends to overlook too much of the interpersonal or social structure of the texts examined. If, however, these two different styles of macroanalysis could be combined, then a richer picture of the combined social and semantic (re)productions of VLSC could be painted.

Thesauri and Conversational Common Sense

If these different sorts of macroanalysis (social network-based and computational linguistics-based) are to be connected together, a linkage point must be found. By looking at the history of discourse analysis, a linkage point can be found: it is the thesaurus.

In the first essay in linguistics to mention discourse analysis, Zelig Harris (1952) provided a rough-draft version of this linkage point. Harris explained that the key to discourse analysis is to find corpus-specific equivalencies:

Suppose our text contains the following four sentences: The trees turn here about the middle of autumn; The trees turn here about the end of October; The first frost comes after the middle of autumn; We start heating after the end of October. Then we may say that *the middle of autumn* and *the end of October* are equivalent because they occur in the same environment (The trees turn here about—), and that this equivalence is carried over into the later two sentences. On that basis, we may say further that *The first frost comes* and *We start heating* occur in equivalent environments. (1952: 6).

In the Anglo-American traditions of discourse analysis, no one has followed up on Harris' work.¹⁴ However, the description provided by Harris on how to find “equivalencies” is a technically unnuanced description of the sort of work that some researchers in contemporary, corpus-based, computational linguistics have undertaken to automatically generate or extend thesauri. Harris's insight about what he called his “distributional analysis of discourse” was that regularities within a given discourse, rendered as “equivalencies,” could be descriptive of the cultural specificities produced and reproduced within a given discourse. By blending the technology of contemporary corpus-based linguistics with Harris's insight, it is possible to use this insight as a pivot point through which different kinds of macro analysis connect together with the concerns of micro analysis of conversation.

Consider the following conversational exchange:

A: What sorts of fruit do you like?

B: Oh, apples and bananas.

A: What about strawberries? Do you buy them when they're in season?

B: No, I don't really like berries.

To find the lexical cohesion between the statements in this exchange, it is necessary to know that apples, bananas, strawberries, and berries in general are all kinds of fruit. Knowing this, it is possible to say that this short sequence concerns fruit, but it is also possible to say that A and B have had a verbal exchange concerning fruit. In short, thesaurus-like knowledge about fruit provides a means for more specifically describing interpersonal and textual relations of the conversation.

Within the micro analysis specialty of conversation analysis, scholars have noted the key role that thesaurus-like knowledge of categories plays in the construction of coherent sequences of dialogue. For instance, in elaborating his theory of categories, the inventor of conversation analysis, Harvey Sacks, provided the following definition and example:

Sacks refers to activities which imply identities as category-bound activities (CBAs). His definition is as follows. *Category-bound activities*: “many activi-

¹⁴ “. . . with Chomsky's appropriation of the notion of transformations as an intrasentential feature, and with the overwhelming dominance of linguistics by the transformational-generative movement which Chomsky came to lead, Harris' early attempt with longer stretches of texts was not followed up, and the models of discourse analysis described below [discourse analysis as influenced by Michael Halliday and conversation analysis as influenced by Harold Garfinkel] cannot be seen as direct developments of Harris's model (Malmkjaer 1991: 100–101). However, Michel Pêcheux and his colleagues and students in France did attempt to use Harris's (or at least Harris-like) insights to examine differences and similarities between specific discourses. See, for instance, Pêcheux (1995).

ties are taken by Members to be done by some particular or several particular categories of Members where the categories are categories from membership categorization devices" (Sacks 1992: 249). CBAs explain why, if the story read "The X cried. The Y picked it up," we might have guessed that X was a baby and Y was a mommy. Crying, after all, is something that babies do and picking up (at least in the possibly sexist 1960s) is something that mothers did. (Silverman 1998: 83)

Similar observations about the key role of semantic and pragmatic associations for given terms in the construction of the coherence and cohesion of texts were realized within some work in computational linguistics (cf. Carbonell 1980). However, this computational linguistics work, like most other non-Harris-like computational work on discourse analysis, has been—for all practical purposes—a methodology of micro analysis of conversation and discourse.¹⁵

What has been left unexplored is the fact that there now exist empirical methods applicable to large-scale corpora that can provide a means for documenting the emergence of categories of terms, what Harris called equivalencies between terms. But, it is not the case that these new techniques from corpus-based linguistics can automatically bridge the theoretical chasms dividing micro from macro conversational analysis and social-network versus computational-linguistic macro analyses. One more theoretical insight is necessary: even as much as the textual and interpersonal relations are influenced by the ideational relations (i.e., the semantic links articulated in thesaurus-like compilations), the inverse is also true. In other words, the social and semantic aspects of VLSC are related in a mutually recursive manner: ideational → textual → interpersonal → ideational →

From the word usages (what Ferdinand de Saussure would call *parole* or what Noam Chomsky would call *performance*) in a corpus of texts, a set of equivalencies and thus a rough idea of semantic relations between terms can be derived with the procedures of corpus-based, computational linguistics. These equivalencies can be compiled as a kind of rough draft thesaurus. The categories and equivalencies in the thesaurus have, in turn, an influence on how cohesion (i.e., textual) and social (i.e., interpersonal) relations are labeled. By looking at which terms are important to a conversation (i.e., which terms label a large number of social and cohesion relations present in a corpus of multiauthored texts, (such as an archive

¹⁵ Most "discourse analysis systems" that have been built in the fields of artificial intelligence and computational linguistics have been very elaborate productions constructed to illustrate the analysis of interchanges that can be transcribed into one or two pages of text (e.g., Allen et al. 1996).

of email messages), one can get a feel for which parts of the rough-draft thesaurus are important. The ways in which these highlighted elements of the rough-draft thesaurus are "spoken about" by members of the conversation provide a means for characterizing the conversation as a whole. Thus, for instance, a conversation that associates water with hydrogen and oxygen might be characteristic about a conversation of chemistry rather than a conversation about alchemy.

As conversations and so cultures and common sense evolve, so do the thesauri that can be derived from them. This is true too of more official, hand-compiled reference works.¹⁶ Very large-scale conversation is an eclectic domain because, as it is currently practiced on the Internet, participants can come from a wide diversity of cultural backgrounds and so what is or is not commonsensical cannot be enumerated beforehand. An understanding of VLSC requires a perspective that allows one to see, for instance over the course of a long-term conversation, how common sense is produced, reproduced, extended, and changed by a group of potentially culturally diverse participants. The political philosopher Antonio Gramsci gives us just such a picture of common sense: "Every social stratum has its own 'common sense' and its own 'good sense,' which are basically the most widespread conception of life and of men. Every philosophical current leaves behind a sedimentation of 'common sense': this is the document of its historical effectiveness. Common sense is not something rigid and immobile, but is continually transforming itself, enriching itself with scientific ideas and with philosophical opinions which have entered ordinary life. . . . Common sense creates the folklore of the future, that is as a relatively rigid phase of popular knowledge at a given place and time" (Gramsci 1971: 326, as cited in Hall 1982: 73).

From this perspective, common sense is accumulated and transformed through the process and productions of science, philosophy, and other powerful conversations, discourses, and practices. This is a perspective that has been useful for understanding the workings of older media (newspapers, television, film) and could be of use to understand and design for new forms of mediation like VLSC.¹⁷

¹⁶ The literary theorist Roland Barthes speaks of the contents of reference books, like thesauri, as "cultural codes" central to the process of reading. "The cultural codes, which are extremely numerous and heterogeneous, to a very large degree subsume all the other categories. They speak the familiar 'truths' of the existing cultural order, repeat what has 'always been already read, seen, done experienced.' . . . Barthes underscores the discursive basis of the 'reality' to which cultural codes refer by equating it with 'the set of seven or eight handbooks accessible to a diligent student in the classical bourgeois educational system'" (Silverman 1983: 241, 274).

¹⁷ According to Stuart Hall (1982), Anglo-American media studies of the early twentieth century saw the media (newspaper, television, etc.) as producers of content that "re-

Maps of Very Large-scale Conversation

Discourse architecture entails two kinds of work: (1) the design and implementation of new computer network technologies for discourse, that is, the means to shape the conversation that takes place within a given system; and (2) the production and employment of analytical techniques for identifying conversational structure and explicating the forces that shape it. This chapter has, so far, discussed only the second kind of work.

I have argued—following Michael Halliday—that conversational common sense has at least three crucial dimensions: the interpersonal, the textual, and the ideational. Using examples from art, comedy, and sociology, I have illustrated how the breakdown of conversational common sense can have effects both comedic and/or dire for social cohesion. I have hinted how new and old thinking about thesauri gives one insight into the constitution of conversational common sense. Finally, I have asserted—with Antonio Gramsci and Stuart Hall—that common sense is an “accretion” dynamically produced and transformed by the groups that it links: conversational common sense is defined in a mutually recursive relationship with the social group that invents and reproduces it.

To illustrate the other aspect of discourse architecture—the design and implementation of new computer network technologies for discourse—I present a system, the Conversation Map, designed to visualize the three dimensions of VLSC common sense and its emergence and transformation. The Conversation Map system accepts a corpus of hundreds or thousands of e-mail messages and analyzes those messages using a set of computational linguistics and sociology techniques to generate a summary of the messages that includes (1) who is reciprocally replying to or quoting from whom—i.e., the interpersonal dimension of the conversation; (2) the themes of discussion that are important to the conversation embodied in the messages—i.e., the textual aspect of the conversation; and (3) what can be understood as some of the emergent definitions or metaphors of the discussion that are apparent if, in a certain sense, all of the participants’ language inscribed in the text—i.e., the content—of the e-mail messages is analyzed and summed together. This last aspect is performed through the automatic calculation of a rough-draft thesaurus from the

lected” the “common sense” of the larger public. The media was said to objectively write down and distribute the consensus, or *sensus communus*, that was produced by the public independent of the media. Hall argues that, later, media studies came to recognize the media’s role in producing, rather than simply reflecting, community values and common sense. By being the only “voice” that could reach across the nation and even across the world, the electronic and print media framed public discourse, and thus public “common sense,” simply through the editorial choice of which stories should be broadcast and which should be left untold.

written content of the e-mail messages. In short, the Conversation Map is designed to make the three dimensions of VLSC common sense visible: the interpersonal, the textual and the ideational. More specific and technical descriptions of the Conversation Map system can be found elsewhere (Sack 2000, 2002). Here we will examine a few examples of the maps that were automatically generated by the system.

One unprecedented activity that the Internet has made possible is the debate about international politics by ordinary citizens in different countries on a daily basis in a public “space” where people do not necessarily know one another before the debate begins. Such debates occur regularly in weblogs, listservs, and Usenet newsgroups. The following examples are all drawn from public, Usenet newsgroup discussions, and they all illustrate ways of understanding these new horizontal, transnational relations conducted by “citizen diplomats.”

The first map (fig. 2) was generated from several hundred messages posted to the Usenet newsgroup soc.culture.palestine during the period August 1–7, 2001. The upper left corner of the map displays a social network. Nodes in the network are message authors. A link between nodes indicates that two authors have mutually responded to and/or quoted from one another. Note the visible evidence of a relatively tight-knit group: there is one large cluster of authors with only a couple of author pairs floating off to the side. The upper center menu lists a series of dis-

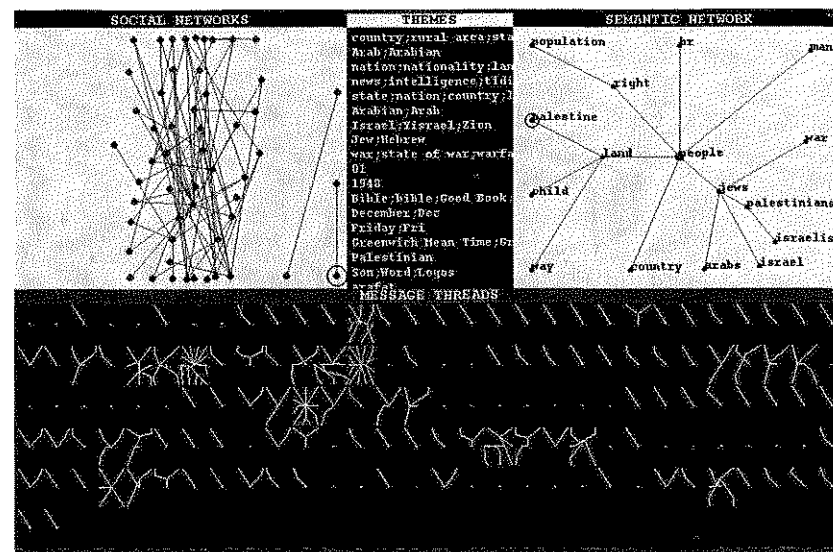


Figure 2. soc.culture.palestine during the period August 1–7, 2001

cussion themes that the Conversation Map has counted in the bodies of serially posted messages (message threads). Note that the terms “Arab and Arabian” are counted as frequent themes of discussion. The upper right corner displays the output of the automatic thesaurus computations: terms that are linked are calculated to be potentially similar terms within the VLSC. Here one can see that “Israelis” and “Palestinians” are counted as possibly similar terms, as are “Jews” and “Arabs” and “lands” and “peoples.”

Each of these three analyses (social network, themes, and semantic network) are cross-linked with one another. This is built into the software so that clicking on one panel highlights terms in the others. Thus, for instance, clicking on a theme highlights that part of the social network in which the participants have discussed the theme. The Conversation Map illustrates how the interpersonal, textual, and ideational aspects of a VLSC are interrelated and furthermore shows how a generated thesaurus reveals some of the crucial equivalences under discussion by the group: How are Palestinians the same as/different from Israelis? What makes Jews like Arabs? To what extent does the definition of a people depend upon a definition of land or country?

The next Conversation Map was generated from the same news group, *soc.culture.palestine*, but messages were collected a few days later; messages analyzed for this map include those posted between August 4 and 11, 2001. In other words, there are some overlaps with the messages analyzed for the map in figure 2. In figure 3, one can see that the terms “Arab” and “Arabian” have ceased to be central themes of discussion. The social network has increased in size and fractured apart into several nonoverlapping clusters, thus indicating that there is not just one conversation taking place but rather several in the same space, that is, the same newsgroup. Note also the possible equivalence drawn between entities common to conventional international relations—the posited equivalences between “government,” “nation,” and “state.” By comparing figure 2s and 3, one can see how the VLSC of the newsgroup *soc.culture.palestine* changed over the course of a week and a half.

Figure 4 is a Conversation Map generated from over one thousand messages posted to the Usenet newsgroup *soc.culture.afghanistan* in the time period September 24–28, 2001. Note the highly fractured social networks. Note also the extremely generic semantic network including only abstract terms like “state,” “country,” “government,” and “people.” Unlike Israel and Palestine, Afghanistan had, at the time, only one Internet service provider. Writers knowledgeable about the specifics of Afghanistan are, consequently, relatively rare online. It is highly unlikely that anyone logged in from Afghanistan to post their side of the story to the newsgroup. In other words, it may be possible that some of these newsgroups

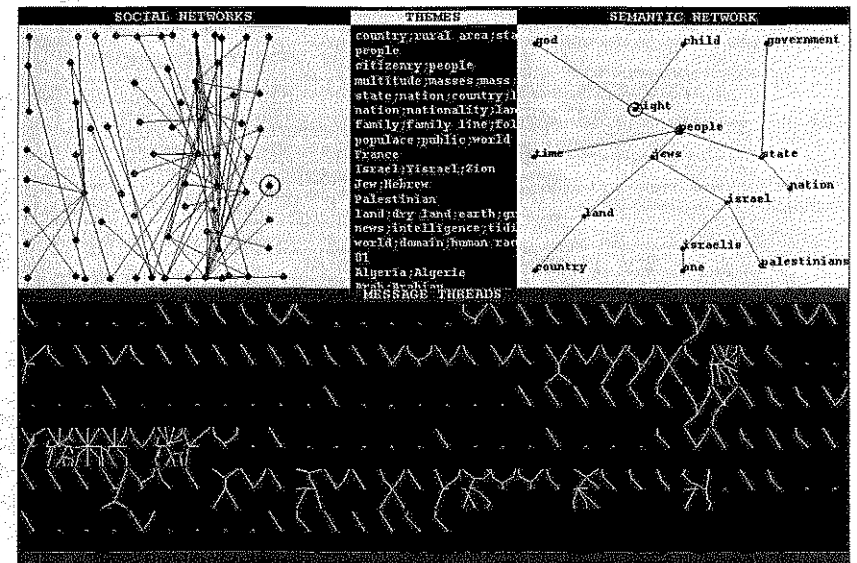


Figure 3. *soc.culture.palestine* during the period August 4–11, 2001

do support a truly new kind of cosmopolitan, citizen-centered diplomacy. But what happens in these cyberspaces is also inflected by what is happening offline. For U.S.-Afghani negotiations to take place in Usenet, the necessary infrastructure for the Internet would have to exist in Afghanistan.

The conversation maps in figures 2, 3, and 4 graphically summarize three collective productions achieved by groups of hundreds of people over the course of several days of online conversation:

1. The social networks shown in the maps give some indication of the interpersonal relations of the groups: they indicate how often and with whom members of the groups are reciprocating messages.
2. The lists of calculated discussion themes are created from a computational analysis of the words that are quoted and repeatedly taken up in sequences of messages. Discussion themes are listed according to the number of participants who have exchanged messages about the theme. Thus, the menus in the figures above can be understood as representations of the intertextual structure of hundreds or thousands of e-mail messages and also as representations of a group's current focus: the themes listed at the top of the list are themes addressed in the messages of many of the participants.
3. Finally, by parsing the contents of all of the messages and recording which words, specifically which nouns, are used in a manner similar to which other

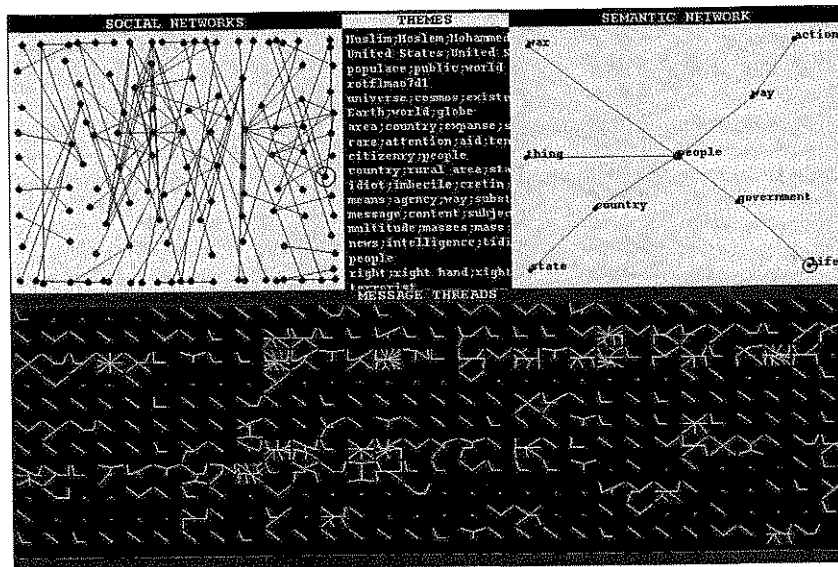


Figure 4. soc.culture.afghanistan during the period September 24–28, 2001

nouns (i.e., which nouns are written about like which other nouns), a rough-draft thesaurus is computed for the group and partially displayed as a semantic network. Two nouns are plotted next to one another in the semantic network if, for instance, many participants have used the same set of adjectives to describe them, have associated the same set of verbs with them, and collocated them in messages alongside a similar set of other nouns. Thus, for example, if in the text of many participants' messages in a given discussion group two words, like "building" and "argument," are both repeatedly described as having foundations (as being solid, strong, shaky, or weak; as collapsing, falling down, or standing up), then they may very well appear next to one another in the semantic network computed by the Conversation Map for that discussion group. This semantic network can be read as symptomatic of the emerging synonyms or metaphors that participants are in the process of collectively creating or defining.

Note, however, that the procedure I have outlined above—keeping track of and comparing the lexical contexts of each noun used in the discussion—does result in a set of many semantic networks. For discussions of the size mapped in the figures above, this results in an analysis of several thousand nouns. Although all of these semantic networks for the thousands of nouns are computed by the Conversation Map, only one semantic network is displayed: the semantic network that contains nouns

frequently used as themes of discussion, and thus the nouns that are arguably the current, collective focus of the group. Interestingly, in all three maps, the noun "country" is within the groups' collective foci.

The maps constitute *representations* of the VLSCs, but they also can be used as *interfaces* into the message archives. Another aspect of the Conversation Map that is not discussed in this chapter is the fact that the maps it outputs are executable as Java applets on the Web. They are therefore in principle accessible to most of the discussion participants since they can be viewed as web pages. By clicking on the various aspects of the maps, one can see how each piece of the map is cross-indexed with other parts of the map. Thus, for instance, clicking on a discussion theme will highlight that portion of the social network that has exchanged messages about the selected theme. (See the main web site for the Conversation Map to learn more about how the maps function as interfaces as well as representations: <http://www.sims.berkeley.edu/~sack/cm>.)

Deliberative Democracy and the New Public Sphere

Is there a politics of very large-scale conversation? There are many such politics, but large, many-to-many exchanges between citizens are of special interest as new forms of deliberative democracy.

The most exciting and potentially revolutionary political application of a computer conferencing system is the facilitation of the direct participation and voting of citizens on important state or national issues. . . . Perhaps the first operational use of computer conferencing systems to facilitate "participatory democracy" will be J. W. Huston's Constitutional Convention project in Hawai'i. Funded by grants from local and mainland foundations, it is being designed to establish 21 community centers throughout the state to allow public participation in the 1978 Hawai'i Constitutional Convention. (Hiltz and Turoff 1994: 195, 197)

For at least a quarter century, many have been excited about the possibilities of computer networks as a means of facilitating democratic participation. Reviewing the area in the mid-1990s, sociologist Manuel Castells (1997: 350) noted that local democracies appeared to be flourishing around the world and that "When electronic means are added to expand participation and consultation by citizens, new technologies contribute to enhanced participation in local government." Collections, such as Tsagarousianou, Tambini and Bryan's *Cyberdemocracy* (1998) documented how these experiments in local, online democracy were progressing in Amsterdam, Athens, Berlin, Bologna, Manchester, Santa Monica, and elsewhere.

At the national level, there has been less interest in citizen-to-citizen communication and more emphasis on delivering government services and documents (tax documents, legal codes) to citizens via the Internet. This understanding of the Internet as a one-way publishing and distribution network rather than as a many-to-many medium is due to a variety of entrenched economic and political interests. It is interesting to compare these political and technical efforts to produce the Internet, especially the Web, as a one-way broadcasting medium with centralized control with similar efforts that have been suffered over the course of the last century, such as the re-creation of radio as a one-way medium (e.g., Neuman, McKnight, and Solomon 1997). In fact, it appears to be the case that—as municipal web sites become more and more common—even local governments seem most intent on supporting a one-way, “services” model of information delivery rather than many-to-many deliberative discussion.

Interesting and powerful exceptions do exist. For example, Stephen Coleman and his colleagues at the Hansard Society have initiated online, public forums to elicit public opinion and encourage democratic deliberation on issues of national importance and communicated these results of online deliberation to the U.K. Parliament.¹⁸

Many of the political scientists and communications experts now exploring the area of online, democratic deliberation have been deeply influenced by philosopher Jürgen Habermas’s (1991) conception of the “public sphere” and its transformations over the past three centuries.¹⁹ Habermas’ contention was that the public sphere constituted a set of norms and forums (the newspaper, the café, etc.) that allowed bourgeois society to meet and, through rational debate and deliberation, find consensus. His diagnosis was dark: rational-critical debate largely disappeared in the twentieth century as citizens became consumers and so consumption—rather than conversation—dominated the forums of the public sphere. However, contemporary work in “community informatics” has proceeded with the hope that computer networks can provide a basis for a renewed public sphere where deliberative democracy can be supported (e.g., Schuler and Day 2000).

Some of the more practical work necessary for the goal of supporting a new, online public sphere is well defined. Exemplary organizations, like the Seattle Community Network (<http://www.scn.org>), provide community members with e-mail accounts; host web sites, online discussion forums, and public calendars; provide help and computer training; and fa-

¹⁸ See Coleman (2004). See also the Hansard Society’s website, <http://www.hansard-society.org.uk/eDemocracy.htm>

¹⁹ See also more recent commentary in Calhoun (1992).

ilitate low-cost or free distribution of computers and other necessary hardware. Some might wonder why this list of technical foundations for a new public sphere does not also include, for instance, multiway video conferencing and/or streaming audio servers for all participants. In principle this would be possible and would allow many citizens to, essentially, run their own television and/or radio stations. But, the practical essence is that most of this technology is too expensive and/or too complicated to support for large groups of people. So community networking of today is especially dependent upon textual exchanges such as e-mail, newsgroups, and weblogs.

Once the technical infrastructure has been put into place, the work that remains is not so well defined: How can online, deliberative discussions be engendered and facilitated? Activists and technologists attempting to support new forms of online democracy have had to turn to philosophy and political theory to help define the crucial issues: What is democracy? What is the “public sphere”? What constitutes deliberative discussion? These seemingly abstract questions have become pressing concerns for community networks. I contend that few, if any, have a working definition of deliberative discussion when the discussion involves asynchronous, online exchanges among hundreds or thousands of people. Even at the local, civic level, online exchanges of this sort quickly reach the size of very large-scale conversations.

Attempts to produce working definitions of new, electronic forms of the public sphere and of large-scale, deliberative discussion can be found in the literatures of the arts, humanities, and social sciences. One body of work is critical insofar as it points out the weaknesses of a Habermasian ideal of the public sphere and its goal of consensus through rational discussion. Habermas’s original focus on the *bourgeois* public sphere was scrutinized, and it has been pointed out that participants in the stated ideal were limited, for instance, by class (e.g., Negt and Kluge 1993), by gender (e.g., Fraser 1992), or by activity—specifically the democratic potential of rational discussion has been questioned by Jean-Francois Lyotard (1984) and others. These critiques have yielded alternative ideals, and alternative ideals for online exchanges have been articulated.²⁰

Another set of work takes as given a specific set of ideals of democratic discourse and then attempts to measure online exchanges against these ideals. This type of work can be problematic if either (1) the stated ideals of democratic discussion recapitulate the weaknesses already scrutinized and critiqued in Habermas’s original work; or (2) the stated claims of the authors exceed the possible reach of their empirical work. The sec-

²⁰ For example, Jodi Dean (2001) points out the limitations of the “public sphere” ideal and prefers the term “civil society.”

ond flaw is caused by a misunderstanding of the scale of online discussion. There are, for instance, many researchers who have closely read a few hundred e-mail messages and now claim to have an evaluation of Usenet as a possible, new form of the public sphere. These sorts of claims are mistaken. To understand the enormity of Usenet, it helps to know that as of late 2003, Google (<http://www.google.com>) had an archive of over eight hundred million messages exchanged on Usenet newsgroups. At best, with an analysis of a few hundred messages, one might claim to have some insights into a fleeting moment of one newsgroup within Usenet. It is with this specific problematic in mind that I propose a theory and technology of very large-scale conversation.

The best of this literature is worth close examination because it points toward many interesting new possibilities. In his book *Democracy in the Digital Age*, Anthony G. Wilhelm examines five hundred messages (fifty messages apiece from six Usenet newsgroups and four AOL discussion groups) and then makes sweeping claims like these: "If a democratic discussion is to be defined at least in part by the quality of the conversation, then the newsgroups analyzed in this study are not very deliberative" (2000: 98).²¹ Examination of fifty messages is unlikely to provide enough evidence to warrant such an evaluation. For example, if one downloads even just the past week's worth of messages from the six Usenet newsgroups examined by Wilhelm, one gets the following message counts: alt.politics.elections (220 messages posted in the past week); alt.politics.libertarian (1,081); alt.politics.media (647); alt.politics.org.cia (104); alt.politics.reform (62); alt.politics.white-power (199).²² Thus, the number of messages per group chosen by Wilhelm does not cover even a given week's worth of messages exchanged on the smallest of the groups.

What is most interesting about Wilhelm's study is his effort to define a set of necessary conditions for deliberative discussion. He then attempts to operationalize those criteria to determine if online discussions are deliberative. Wilhelm's criteria of deliberative discussion are stated as a series of research questions:

1. **Reciprocation:** . . . To what extent do participants of virtual groups solely provide ideas and information versus seeking information from other forum members? . . . [Do] reciprocal acts occur . . . in which participants . . . articulate their interests through talking, sharing ideas, and negotiating differences[?] (88)
2. **Interactivity:** . . . To what extent do participants of political groups ex-

²¹ For an earlier version of Wilhelm's work in which he arrives at even more sweeping conclusions, see Wilhelm (1999: 154–78).

²² Messages from these groups were downloaded on October 12, 2003, from the Usenet news server news.ucsc.edu.

change opinions as well as incorporate and respond to others' viewpoints? (88–89)

3. **Heterogeneity:** . . . To what extent is there in-group homogeneity of political opinion on Usenet newsgroups? . . . In this case, homogeneity is defined as the extent to which individual messages adhere to a certain political affiliation. (89).

4. **Rationality:** . . . To what extent are substantive, practical questions debated rationally in contradistinction to *ad hominem* argumentation not susceptible to criticism and grounding? (90)

Insofar as Wilhelm is one of many researchers inspired by Habermas, his criteria are representative of a larger literature on the issue of online, deliberative democracy. Moreover, his suggestions for operationalizing these criteria into a means of empirically investigating these questions suffer from the same problem as other empirical work in the literature: the work does not scale to the size necessary to address the questions posed about public, online discussions like those of Usenet newsgroups. Wilhelm attempts to address these questions with a form of content analysis²³ that requires a panel of judges to comb through and categorize the messages: a very laborious process that would cost a fortune to apply to thousands of messages. Other tested methodologies—for example, survey methods—suffer from the same problem.²⁴

To adequately address the questions posed by Wilhelm regarding Usenet newsgroups and other VLSC, it is necessary to have a theory and methodology of VLSC that is at least partially embodied in a piece of software. This genre of "theoretical software" has been, historically, well known in social science but is recently less common than it used to be.²⁵

The Conversation Map simultaneously embodies and articulates a theory of VLSC and allows one to begin to address the questions posed by Wilhelm. Consider, for example, a Conversation Map of one week's worth of messages posted to one of the Usenet newsgroups studied by Wilhelm, alt.politics.elections. The map shown in figure 5 was calculated from over a thousand messages posted to the group in the week before the 2000 U.S. presidential election. Illustrated below are examples of how

²³ For a definition of content analysis, see Krippendorff (1980).

²⁴ See, for example, the survey methods employed by Cappella, Price, and Nir (2002: 73–93). Although these researchers were able to survey a large number of people, their methodology entailed the creation of a set of small (approximately 20 to 30 people), moderated, nonpublic discussion groups that ran periodically for one hour. Thus, it is unclear whether their results have anything to say about VLSC: large, online, ongoing, public, unmoderated discussions involving hundreds or thousands of people at once.

²⁵ See, for example, Robert Abelson's early, computational/theoretical analyses of ideology, belief, and opinion that were embodied in working simulations; e.g., his "Goldwater Machine" (Abelson 1973).

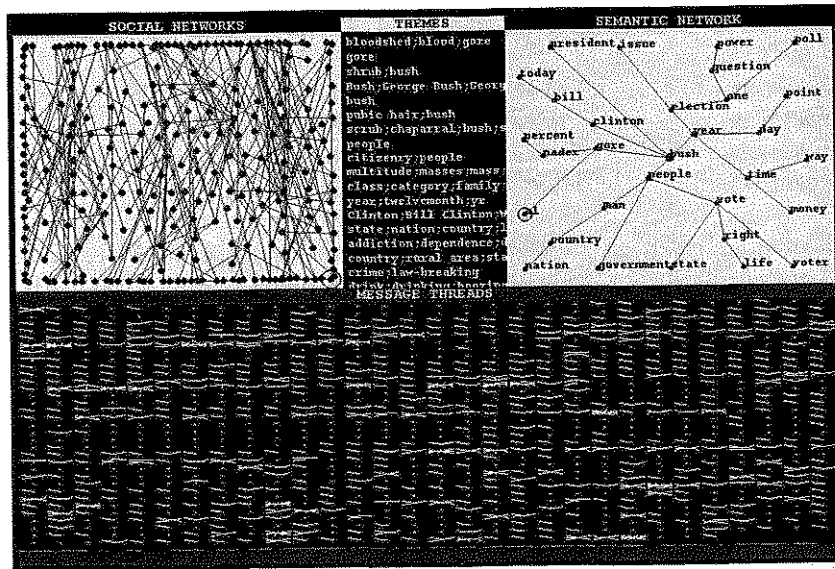


Figure 5. alt.politics.election for the week prior to the 2000 U.S. election

one might begin to address Wilhelm's questions with maps calculated by the Conversation Map system.

1. **Reciprocation:** Are participants reciprocating with one another; that is, are they responding to and/or quoting from the messages of other participants? The social networks provide a partial representation with which one can explore this question. In the map shown in figure 5, the answer to this question is not a simple "yes" or "no." Here it is possible to see a great number of social networks (recall that two participants are connected in the network if they have replied to each other and/or quoted from one another). Compare the social network shown here with the one shown in figure 2, where practically all the visible participants are integrated into a single network.

2. **Interactivity:** The Conversation Map graphically shows two ways in which participants are (or are not) incorporating or responding to others' postings.

2a. The calculated "themes of discussion" indicate which topics were repeatedly addressed in sequences of messages exchanged. These sequences are normally termed "message threads." The lower half of every Conversation Map is a graphical representation of all of message threads analyzed. By clicking on a given theme in the menu of themes, one can see which threads address the selected theme. The ovals highlighting the lower half of the Conversation Map shown in figure 6 indicate all of the messages threads where then-Vice President Gore was a theme of discussion.

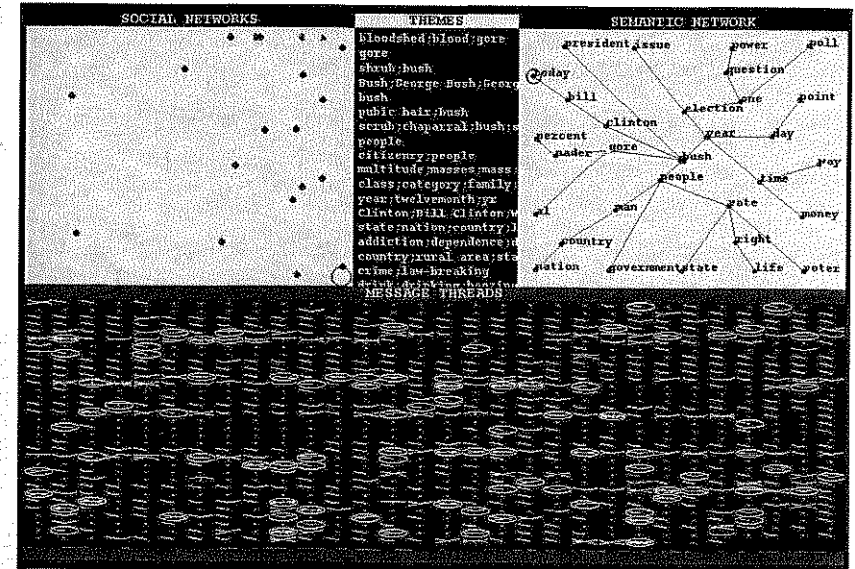


Figure 6. The same conversation map as shown in figure 5 with the discussion theme "Gore" selected

2b. The portion of the social network shown in figure 6 contains those pairs of discussants who have exchanged messages on the theme of Vice President Gore.

Well-focused, interactive newsgroups tend to have several themes of discussion that are repeatedly addressed in multiple message threads and cover substantial portions of the social network. In a sense, these two graphics show the extent to which a group of discussants stays on topic.

3. **Heterogeneity:** When "summed together," does the language of the messages exchanged link together a diverse semantic field, or is it a relatively homogeneous set of comments? Examination of the calculated semantic networks (through double-clicking on terms and sets of terms in the semantic network) reveals the diversity of terms employed to describe the themes of discussion. For example, after selecting both "Bush" and "Gore" in the semantic network, one can double click to demand a list of the terms used in the text of messages to describe Bush, the terms used to describe Gore, and the terms that were applied to both Gore and Bush (see fig. 7). In the interface, the different lists of terms are distinguished by color.

Symptomatic of the heterogeneity of opinion for a given topic is the diversity of lexical terms associated with the topic. For example, from the (partial) list of verbs displayed in figure 7, one can see that Bush was the

ASSOCIATIONS FOR BUSH AND GORE

ACTIONS DONE BY BUSH OR GORE abandon accomplish accuavlate
 accuse achieve acknowledge add admit allow announce answer anticipate
 appoint argue arrest arrive ask assure attack attend authorize avoid
 base be become begin believe belittle biased blow break bring build
 buy call cancel capture care carry catch characterize choose cite
 claim clash close come comment comply compromise conceal concede
 congratulate contend continue control convict cooperate count create
 criticize cut deal deceive deciaate declare defeat defend define
 demand demonstrate deny deserve design designate destroy disclose
 discredit dislike divide do draw drink drive drop elect exhrace end
 enjoy enlist enroll ensure enter equate erupt evade exaggerate excuse
 exhale explain extend fail fall favour fight find finish fire fix
 flunk fly focus fora fortify fuable gain get give go handle hang
 happen hate have he hear help hide hike hold hurt implemt implicat
 impose impress increase insist invest involve issue join keep kick
 kill know lambaste lead leave let lie like live look lose maintain
 make marry match meet miss move need nominate note observe offer
 oppose order overcoe overturn ove parse pass pay persist phone pit
 place play plead please plow possess predict prefer pretend prevail
 prioritized profit project promise propose protect prove
 provide pull push put quadruple qualify question quit raise rally
 re\present realize receive recite recognize recommend reconcile
 refer relate release rely remain remember reply report represent
 respond restore retract return reveal reward rin run running runs
 save saw say scend scrub see seen seamed sell send serve set settle
 sever sev share shift shoot show sign sing sit snear sound speak
 spell spend sponsor stage stand star start startle state stick stop
 stretch strike stumble suck suckle suggest summarize support sway

Figure 7. A partial list of the terms associated with Bush and/or Gore

subject of the following verbs in messages posted to the group: Bush acknowledges, adds, admits, announces, authorizes, avoids, belittles, compromises, conceals, drinks, promises, reveals, suggests, sways, etc. Clicking on any one of these terms causes the Conversation Map to produce a hyperlinked list of sentences in which the term appears. Figure 8 shows such a hyperlinked sentence generated by clicking on the verb “drink.” Clicking on a sentence allows one to examine the message in which the sentence appears.

4. **Rationality:** Although Wilhelm is interested in evaluating whether or not a group is debating questions “rationally” (according to criteria of knowledge, truth, and conditions of validity outlined by Habermas) he—like many others interested in these criteria (e.g., Cappella, Price, and Nit 2002) has had to evaluate “rationality” according to the structure of the arguments advanced and the number of reasons included to substantiate an argument. In short, what is empirically decidable is not the rationality of argumentation, but rather the rhetorical structure of the online exchanges. This should make sense because rhetorical arguments tend to be well structured. Some of this information is avail-

SENTENCES WITH BUSH	
DRINK : ACTION DONE BY BUSH	
PARTICIPANT	SENTENCE
	• Bush can not drink socially because he would fall off the wagon.

Figure 8. A sentence from the newsgroup associating a term (Bush) with a verb

able through (a) a close examination of the characteristic message thread structures of a given group, some is visible in (b) the quoting and citation patterns manifest in the messages, and some could be calculated automatically given (c) a procedure for rhetorical structure parsing (e.g., Marcu 1997). The Conversation Map incorporates a means for computing (a) and (b) and displays those results in thread structures seen in the lower half of each Conversation Map (which can be further magnified and explored by clicking on each thread, as has been done for one thread in figure 9), and in the social network and the messages themselves (in which quotes are identified automatically and hyperlinked to the message(s) of origin). A partial implementation of (c) is imple-

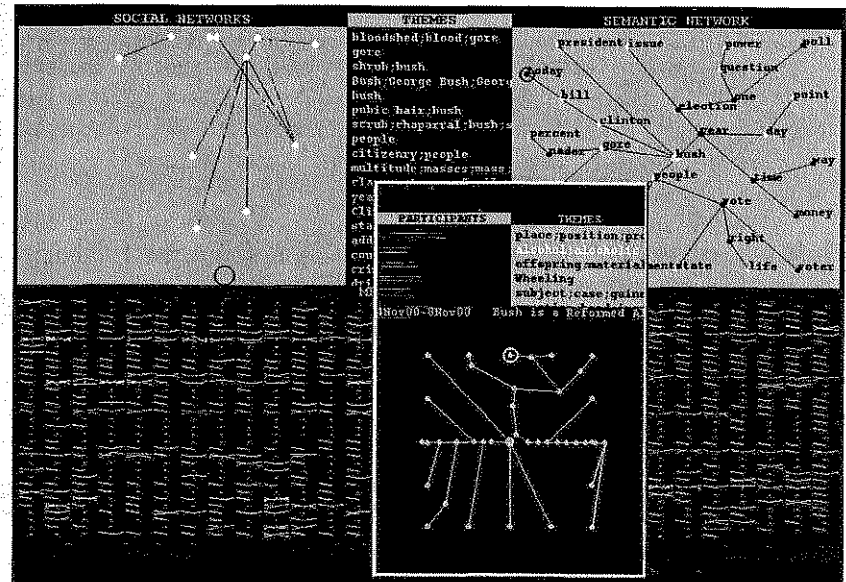


Figure 9. Close examination of the structure of a thread

mented in the Conversation Map: discourse markers indicative of structured argumentation (“because,” “therefore,” etc.) are tagged; however, the results of this tagging process are not currently visible within the Conversation Map interface.

Examination of the pattern displayed in the lower half of a Conversation Map gives some idea of characteristic length and structure of the threads in an online discussion. The Conversation Map plots the threads in a radial, spider web pattern: the initial message is located in the middle of the window, responses to the message are plotted in a circle around the middle and connected to the middle, responses to response are plotted in a ring slightly larger than that, and so forth. If each message had exactly the same number of responses, then the thread would look like a spider web. The more common case is a thread with many asymmetries, like the one shown in figure 9: messages differ widely according to their number of responses. Note that threads containing only one message (and no responses) appear as a dot; threads with one post and one response appear as a line. Figure 9 reveals that many of the posts to the group are unrequited and/or garnered only one response. Compare this to the alternations of simple and elaborately structured threads that one can see in figures 2, 3, and 4.

While the Conversation Map does not answer all of the questions posed by Wilhelm concerning the deliberative quality of discussion in a newsgroup, it does provide a means for exploring those questions because it incorporates a theory of VLSC into a computerized analysis procedure and a graphical interface. Consequently, it is possible to see how variable and quickly changing a discussion group can be with respect to any of these criteria. For instance, figure 10 shows a map of the same discussion group (alt.politics.elections) for the week immediately following the U.S. presidential election of 2000.

Note, among other things, how one can see in figure 10 that the conversation has shifted away from a conversation about the candidates (Bush, Gore, Nader) and is now a discussion about the technicalities of the election: vote, count, ballot, election, and so on are the central themes of discussion. If nothing else, this map illustrates how quickly a discussion can change, and therefore how careful one should be about generalizing from a one-time analysis of a newsgroup.

These Conversation Map images also illustrate the utility of the theoretical topography used to visually investigate online discussions and which I have argued for throughout this chapter. The topography of discussion proposed here is one of links, associations, graphs, and networks. The topography of a Habermasian democratic exchange is one of idealized spaces, territories, and logics—like the so-called public sphere and

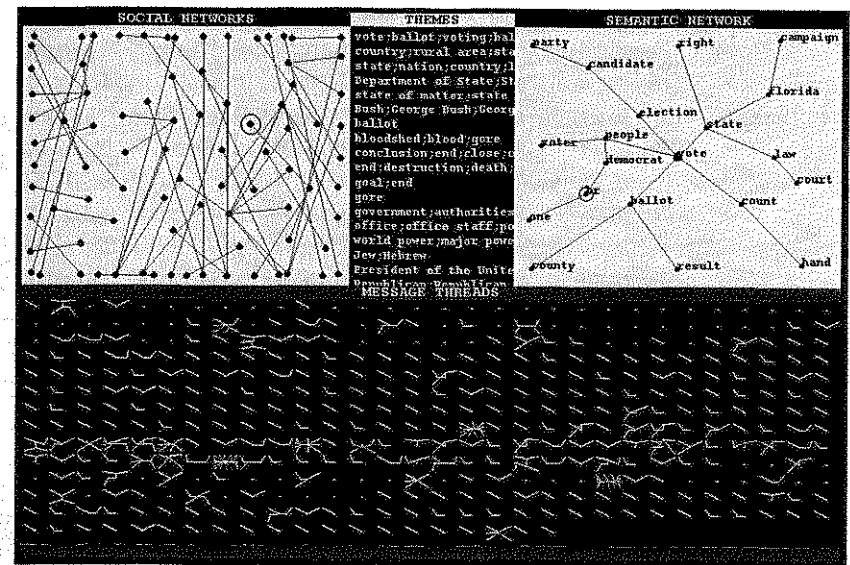


Figure 10. alt.politics.election for the week after the 2000 U.S. election

rationality. As the following quote shows, this difference in theoretical vocabularies—that based on idealized spaces, territories, areas, and logics versus that based on links, associations, and networks—is a centuries-old dispute. In his comments on sixteenth-century political theorist Guillaume de la Perrière (La Perrière 1567), Michel Foucault (1991: 93) argues that

what government has to do with is not territory but rather a sort of complex composed of men and things. The things with which in this sense government is to be concerned are in fact men, but men in their relations, their links, their imbrication with those other things which are wealth, resources, means of subsistence, the territory with its specific qualities, climate, irrigation, fertility, etc.; men in their relation to that other kind of things, accidents and misfortunes such as famine, epidemics, death, etc.

While Foucault's words help to explain the longer genealogy of political analysis that the Conversation Map takes part in (and which also serves to differentiate it from the Habermas-inspired work of Wilhelm and others), they also point out the many inadequacies in links and networks displayed by the Conversation Map. The links visible in the Conversation Map are only those between people, between words, and between words and people. The larger set of possible relations between people and things

is not visible. This larger set crucial to any expanded understanding of democracy and governance is termed a "Parliament of Things" by Bruno Latour (1993: 144): "we do not have to create this Parliament out of whole cloth, . . . [w]e simply have to ratify what we have always done. . . . Half of our politics is constructed in science and technology. The other half of Nature is constructed in societies. Let us patch the two back together, and the political task can begin again."

These insights from Foucault and Latour provide a direction for future development of the Conversation Map and a practical means for theorizing and analyzing online, deliberative, democratic, very large-scale conversations.

Conclusions: Technologies of the Self and Design Ethics for VLSC

The majority of this chapter has been devoted to what might be called an epistemological inquiry into VLSC. I have attempted to show what kinds of knowledge are necessary to understand these large-scale, online discussions. I have articulated a theory of VLSC and compared it to related work in political science, sociology, linguistics, and philosophy. And I have incorporated this theory into a technology, the Conversation Map, intended to graph the shapes and forms of online discussions. Yet, what kind of technology is the Conversation Map and other work comparable to it?

Michel Foucault distinguished techniques and *technologies of the self* from *technologies of power*.²⁶ For instance, a practice that makes good sense to do for oneself—for example, seeing a doctor on a regular basis and keeping detailed records of one's health—can shift from being a technology of the self to becoming a technology of power if a third party—like an insurance company—is allowed to collect and analyze health

²⁶ "My objective for more than twenty-five years has been to sketch out a history of the different ways in our culture that humans develop knowledge about themselves: economics, biology, psychiatry, medicine, and penology. The main point is not to accept this knowledge at face value but to analyze these so-called techniques that human beings use to understand themselves.

As a context, we must understand that there are four major types of these "technologies," each a matrix of practical reason: (1) technologies of production, which permit us to produce, transform, or manipulate things; (2) technologies of sign systems, which permit us to use signs, meanings, symbols, or signification; (3) technologies of power, which determine the conduct of individuals and submit them to certain ends or domination, an objectivizing of the subject; (4) technologies of the self, which permit individuals to effect by their own means, or with the help of others, a certain number of operations on their own bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality. (Foucault 1997: 224–25)

records. However, a technology of the self can be designed in such a way that makes it more resistant against such a transformation. Thus, it is a good idea to encrypt medical records stored in computer databases and design the database system so that any third parties must request the permission of the patient to get the "key" to records (e.g., Rind et al. 1997: 138–41).

The imperatives of design that shape technologies like the Conversation Map are not strictly epistemological in form. The designer of such technologies and techniques must instrumentalize a set of ethical considerations that make them either, more clearly, a technology of power or a technology of the self. Especially tools of democracy should be designed to be technologies of the self.

There is a long history of the use of media as technologies of the self, as reflective and communicative media for the construction of social, psychological, economic, and political self-governing people and peoples. Diaries have been used for millennia by particular people as a medium for self-reflection, for writing down and shaping the person's image of self. The diary is a medium that functions as a technology of the self where "self" is understood to be the self of one person. The oral story telling practices of folktales function in an analogous manner for the formation and description of a slightly larger self, a self of a small group of people. Oral story telling of folktales is a means for articulating the values and identity of small, tightly knit clusters of people. The facilitation of the production of larger selves, of the selves of self-governing nations, for instance, requires a different kind of medium. Scholars have shown how the mass production capabilities of high-speed printing presses made possible the media of novels and newspapers that were essential to the formation of the modern nation-state (Anderson 1983). VLSCs do and can function as the substrate for new kinds of selves, new sorts of groups of people, that are as yet unnamed (Deleuze and Guattari 1988: 469–71). These new groups of people can be transnational or international in scope. The Conversation Map is intended to be a technology of the self for VLSC.

But the metrics and graphics computed by the Conversation Map could be mishandled as technologies of power. Consideration of the history of "social metrics" and their graphics is sobering, as Ian Hacking points out in his article "How Should We Do the History of Statistics?" (1991: 181): "Statistics has helped determine the form of laws about society and the character of social facts. . . . Moreover the collection of statistics has created, at the least, a great bureaucratic machinery. It may think of itself as providing only information, but it is itself part of the technology of power in a modern state." In short, statistics has been and is "state-istics"—a technology of that not necessarily democratically governed political form that we know today as the nation-state. As VLSC flows over and across the boundaries of today's nation-states, new, as-yet-unnamed political

formations are created through text and talk. These new non-nation-state entities demand new forms of representation that exceed the statistics of the state. Participants within VLSC-based groups need not only representation but also orientation: maps, charts, interfaces, and instruments of navigation in order to locate political position and agency. If these instruments of navigation are to be for the people of democratic organization, then they cannot be—as statistics now is—a tool for only specialists and powerful “decision makers.”

The Conversation Map has been designed to make it difficult to find and follow any given individual author. It has also been designed to output qualitative diagrams rather than quantitative summaries. In these ways the Conversation Map has been made “surveillance resistant.” Furthermore, the “output” of the Conversation Map is a format that can be widely distributed on the Web (a Java applet that runs as an interface to the archive of messages) and thus is, in principle, accessible to any participant of an online discussion. The maps generated by the system are intended to be representative and evocative enough to allow them to function as a means of reflection for the group of people involved in the VLSC—that is, to function as a technology of the self where the self in question is the collective of discussants. But the maps are also intended to be abstract and vague enough to be difficult to use as a technology of power. It is a challenge to find a form of representation that works as a technology of the self but does not work as a technology of power. Others in the social sciences are attempting to find and/or develop such forms of representation, notably ethnographers who write for the benefit of their informants instead of, or in addition to, their fellow anthropologists (e.g., Clifford and Marcus 1986).

Obviously the calculations and interface components of the Conversation Map are far from being universally accessible, and so what is outlined above is simply the first step in a long search for democratic representations and interfaces for VLSC. This search for new forms of self-representation is intrinsic to the development of an ethics of discourse architecture, that is, design for the medium of VLSC. As Guattari states (1991: 2), “In addition to ecology, the question of ethics of media and the future direction of new communication technologies of artificial intelligence and command-and-control constitutes one of the two axes in which to rethink the idea of progress for today’s planet.”

References

Abelson, Robert P. 1973. The Structure of Belief Systems. In *Computer Models of Thought and Language*, edited by R. C. Schank and K. M. Colby. San Francisco: W. H. Freeman and Company.

- Allen, James F. et al. 1996. A Robust System for Natural Spoken Dialogue. In the *Proceedings of the 34th Meeting of the Association for Computational Linguistics*, Santa Cruz, CA.
- Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Bagga, Amit, Breck Baldwin, and Sara Shelton, eds. 1999. *Coreference and Its Applications, Proceedings of the Workshop*. New Brunswick, NJ: Association for Computational Linguistics, June 22.
- Calhoun, Craig, ed. 1992. *Habermas and the Public Sphere*. Cambridge: MIT Press.
- Callon, Michel, John Law, and Arie Rip, eds. 1986. *Mapping the Dynamics of Science and Technology*. London: Macmillan.
- Cappella, Joseph N., Vincent Price, and Lilach Nir. 2002. Argument Repertoire as a Reliable and Valid Measure of Opinion Quality: Electronic Dialogue during Campaign 2000. *Political Communication* 19:73–93.
- Carbonell, Jaime G. 1980. Towards a Process Model of Human Personality Traits. *Artificial Intelligence* 15:49–74.
- Carroll, Lewis. 1960. Alice’s Adventures in Wonderland. In *The Annotated Alice*. New York: Clarkson N. Potter.
- Castells, Manuel. 1997. *The Information Age: Economy, Society and Culture*. Vol. 2: *The Power of Identity*. Oxford: Blackwell.
- Clifford, James, and George E. Marcus. 1986. *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley: University of California Press.
- Coleman, Stephen. 2004. Connecting Parliament to the Public via the Internet: Two Case Studies of Online Consultations. *Information Communication & Society* 7(1)(March):1–22.
- Dean, Jodi. 2001. Cybersalons and Civil Society: Rethinking the Public Sphere in Transnational Technoculture. *Public Culture* 13(2): 243–65.
- Deleuze, Gilles, and Félix Guattari. 1988. *A Thousand Plateaus: Capitalism and Schizophrenia*, translated by Brian Massumi. London: Athlone Press.
- Erickson, Tom, Susan Herring, and Warren Sack. 2002. Workshop Description: Discourse Architectures: Designing and Visualizing Computer-Mediated Conversation. In *Proceedings of ACM CHI 2002*. Minneapolis, April 22.
- Foucault, Michel. 1991. Governmentality. In *The Foucault Effect: Studies in Governmentality*, edited by Graham Burchell, Colin Gordon and Peter Miller. Chicago: University of Chicago Press.
- . 1997. Technologies of the Self. In *Ethics: Subjectivity and Truth; Essential Works of Foucault 1954–1984*. Vol. 1. ed. by Paul Rabinow. New York: The New Press.
- Fraser, Nancy. 1992. Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In *Habermas and the Public Sphere*, edited by Craig Calhoun, 109–42. Cambridge: MIT Press.
- Freud, Sigmund. 1960. *Jokes and Their Relation to the Unconscious*, translated and edited by James Strachey. New York: Norton.
- Garfield, Eugene. 1979. *Citation Indexing: Its Theory and Applications in Science, Technology and Humanities*. New York: John Wiley.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Cambridge: Blackwell.

- Gramsci, Antonio. 1971. *Selections from the Prison Notebooks*. London: Lawrence and Wishart.
- Green, Stephen. 1997. Automatically Generating Hypertext by Computing Semantic Similarity. Ph.D. diss., University of Toronto.
- Grefenstette, Gregory. 1994. *Explorations in Automatic Thesaurus Discovery*. Boston: Kluwer Academic Publishers.
- Guattari, Félix. 1991. Pour une éthique des médias. *Le Monde*, November 6: 2.
- Gumperz, John J., Gurinder Aulakh, and Hannah Kaltman. 1982. Thematic Structure and Progression in Discourse. In *Language and Social Identity*, edited by John J. Gumperz. Cambridge: Cambridge University Press.
- Habermas, Jürgen. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge: MIT Press.
- Hacking, Ian. 1991. How Should We do the History of Statistics? In *The Foucault Effect: Studies in Governmentality*, edited by Graham Burchell, Colin Gordon and Peter Miller. Chicago: University of Chicago Press.
- Hall, Stuart. 1982. The Rediscovery of "Ideology": Return of the Repressed in Media Studies. In *Culture, Society, and the Media*, edited by Michael Gurevitch et al. New York: Routledge.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar*. 2d edition. London: Edward Arnold.
- Harabagiu, Sanda, and Dan Moldovan. 1999. Enriching the WordNet Taxonomy with Contextual Knowledge Acquired from Text. In *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, edited by S. Shapiro and L. Iwanska. Cambridge: AAAI/MIT Press.
- Harris, Zelig. 1952. Discourse Analysis. *Language* 28:1-30, and 474-94.
- Hearst, Marti. 1998. Automated Discovery of WordNet Relations. In *WordNet: An Electronic Lexical Database*, edited by Christiane Fellbaum. Cambridge: MIT Press.
- Hiltz, Starr Roxanne, and Murray Turoff. 1994 (1978). *The Network Nation: Human Communication via Computer*. Rev. edition. Cambridge: MIT Press; first edition: New York: Addison-Wesley.
- Ionesco, Eugene. 1958. *The Bald Soprano and Other Plays*, translated by Donald M. Allen. New York: Grove Press.
- Jakobsen, Roman. 1985. *Verbal Art, Verbal Sign, Verbal Time*, edited by Krystyna Pomorska and Stephen Rudy. Oxford: Blackwell.
- Krippendorff, Klaus. 1980. *Content Analysis: An Introduction to Its Methodology*. New York: Sage Publications.
- La Perrière, Guillaume de. 1567. *Le miroir politique, contenant diverses manieres de gouverner & policer les republicques, qui sont, & ont esté par cy deuant: occure, non moins utile que necessaire à tous monarches; rois, princes, seigneurs, magistrats & autres qui ont charge du gouvernement ou administration d'icelles*. Paris: Pur V. Norment, & I. Bruneau.
- Latour, Bruno. 1993. *We Have Never Been Modern*. Cambridge: Harvard University Press.
- Law, John, and John Hassard, eds. 1999. *Actor Network Theory and After*. Oxford: Blackwell/The Sociological Review.

- Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.
- Liotard, Jean-Francois. 1984. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.
- McHoul, A. 1994. Discourse. In *The Encyclopedia of Language and Linguistics*, edited by R. E. Asher et al., 944-45. New York: Pergamon Press.
- Magerman, D. 1994. Natural Language Parsing as Statistical Pattern Matching. Ph.D. diss., Stanford University.
- Malmkjaer, Kirsten. 1991. Discourse and Conversational Analysis. In *The Linguistics Encyclopedia*, edited by Kirsten Malmkjaer. New York: Routledge.
- Marcu, Daniel. 1997. The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts. Ph.D. diss., University of Toronto.
- Munro, Alan, Kristina Hook and David Benyon, eds. 1999. *Social Navigation of Information Space*. New York: Springer Verlag.
- Negt, Oskar, and Alexander Kluge. 1993. *Public Sphere and Experience: Toward an Analysis of the Bourgeois and Proletarian Public Sphere*. Minneapolis: University of Minnesota Press.
- Neuman, W. Russell, Lee McKnight, and Richard Solomon. 1997. *The Gordian Knot: Political Gridlock on the Information Highway*. Cambridge: MIT Press.
- Pêcheux, Michel. 1995. *Automatic Discourse Analysis*, edited by Tony Hak and Niels Helsloot, translated by David Macey. Amsterdam: Rodopi.
- Pennycook, Alastair. 1994. Incommensurable Discourses. *Journal of Applied Linguistics* 15(2).
- Rind, D. M., et al. 1997. Maintaining the Confidentiality of Medical Records Shared over the Internet and World Wide Web. *Annals in Internal Medicine* 127(2):138-41.
- Sack, Warren. 2000. Conversation Map: An Interface for Very Large-Scale Conversations. *Journal of Management Information Systems* 17(3) (Winter): 73-92.
- . 2002. What Does a Very Large-Scale Conversation Look Like? *Leonardo: Journal of Electronic Art and Culture* 35(4) (August):
- Sacks, Harvey. 1992. *Lectures on Conversation. Vols. 1 and 2*, edited by Gail Jefferson. Cambridge: Blackwell.
- Schiffirin, Deborah. 1994. *Approaches to Discourse*. Cambridge: Blackwell.
- Schuler, Doug, and Peter Day, eds. 2000. *Proceedings of DIAC 2000: Shaping the Network Society: The Future of the Public Sphere in Cyberspace*. Seattle: Computer Professionals for Social Responsibility.
- Silverman, David. 1998. *Harvey Sacks: Social Science and Conversation Analysis*. New York: Oxford University Press.
- Silverman, Kaja. 1983. *The Subject of Semiotics*. New York: Oxford University Press.
- Sinclair, J. M., and M. Coulthard. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. London: Oxford University Press.
- Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville.
- Teil, Geneviève, and Bruno Latour. 1995. The Hume Machine: Can Association Networks Do More than Formal Rules? In *Stanford Humanities Review* 4(2): 47-65.

- Tsagarousianou, Roza, Damian Tambini, and Cathy Bryan. 1998. *Cyberdemocracy: Technology, Cities and Civic Networks*. New York: Routledge.
- Turner, Bryan S. 1996. The Micro-Macro Problem. In *The Blackwell Companion to Social Theory*, edited by Bryan S. Turner. Malden, MA: Blackwell.
- Wasserman, Stanley, and Joseph Galaskiewicz, eds. 1994. *Advances in Social Network Analysis: Research in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage Publications.
- Whittaker, Steve, et al. 1998. The Dynamics of Massive Interaction. In *Proceedings of the International Conference on Computer-Supported Cooperative Work*, 257–64. Seattle: Association for Computing Machinery.
- Wilhelm, Anthony G. 1999. Virtual Sounding Boards: How Deliberative Is Online Political Discussion? In *Digital Democracy: Discourse and Decision Making in the Information Age*, edited by Barry N. Hague and Brian D. Loader. New York: Routledge.
- . 2000. *Democracy in the Digital Age: Challenges to Political Life in Cyberspace*. New York: Routledge.
- Ziff, P. 1960. *Semantic Analysis*. Ithaca: Cornell University Press.

Transnational Communication and the European Demos

LARS-ERIK CEDERMAN AND PETER A. KRAUS

Digitization is triggering a global communication revolution that has the potential to constitute new social domains. Emerging at the intersection of technological and societal processes, digital formations are creating new political topologies and reconfiguring existing networks and organizations. The European Union (EU) is often thought of as such a border-transcending communicative space. Since the heady pioneering days of Jean Monnet, technological advances have inspired the architects of the European integration project. Today, information technology plays a prominent role in the debate about how to promote a closer union of Europe's peoples.

In this chapter, we focus on the question whether digitization could constitute a new political realm coinciding with the European Union. Can democracy take root at the European level without a "demos" encompassing the whole of the Union? How would such a popular unit have to be constituted in order to form a communicative space supporting democratic deliberations and decisions? More specifically, what are the prospects of a digital demos in Europe?

Arguing that democracy does not depend on the preexistence of an ethnic nation, Jürgen Habermas insists that the formation of a postnational polity at the European level is both desirable and possible. Inspired by Habermas's theory of communicative action, many analysts and practitioners agree that the EU constitutes a nascent public sphere offering new opportunities for democratic participation in areas that were previously beyond the reach of national governments. In their view, revolutionary advances in information technology, including the emergence of the Internet, are currently laying the foundations for a viable communicative structure spanning across Europe's national borders.

We would like to thank the anonymous reviewers, as well as Cathleen Kantner, Ruud Koopmans, Dieter Rucht, Tobias Theiler, and James Tully for their useful comments.