Computer-Mediated Discourse Analysis:
An Approach to Researching Online Behavior

Susan C. Herring

School of Library and Information Science
Indiana University
herring@indiana.edu

Introduction

Over the past fifteen years, the Internet has triggered a boom in research on human behavior. As growing numbers of people interact on a regular basis in chat rooms, Web forums, listservs, email, instant messaging environments and the like, social scientists, marketers and educators look to their behavior in an effort to understand the nature of computer-mediated communication and how it can be optimized in specific contexts of use. This effort is facilitated by the fact that people engage in socially meaningful activities online in a way that typically leaves a textual trace, making the interactions more accessible to scrutiny and reflection than is the case in ephemeral spoken communication, and enabling researchers to employ empirical, micro-level methods to shed light on macro-level phenomena.

Despite this potential, much research on online behavior is anecdotal and speculative, rather than empirically grounded. Moreover, Internet research often suffers from a premature impulse to label online phenomena in broad terms, e.g., all groups of people interacting online are "communities";[1] the language of the Internet is a single style or "genre".[2] Notions such as "community" and "genre" are familiar and evocative, yet notoriously slippery, and unhelpful (or worse) if applied indiscriminately. An important challenge facing Internet researchers is thus how to identify and describe online phenomena in culturally meaningful terms, while at the same time grounding their distinctions in empirically observable behavior.

Online interaction overwhelmingly takes place by means of discourse. That is, participants interact by means of verbal language, usually typed on a keyboard and read as text on a computer screen. It is possible to lose sight of this fundamental fact at times, given the complex behaviors people engage in on the Internet, from forming interpersonal relationships (Baker, 1998) to implementing systems of group governance (Dibbell, 1993; Kolko & Reid, 1998). Yet these behaviors are constituted through and by means of discourse: language is doing, in the truest performative sense, on the Internet, where physical bodies (and their actions) are technically lacking (Kolko, 1995).

Of course, many online relationships also have an offline component, and as computer-mediated communication becomes increasingly multimodal, semiotic systems in addition to text are becoming available for conveying meaning and "doing things" online (cf. Austin, 1962). Nonetheless, textual communication remains an important online activity, one that seems destined to continue for the foreseeable future. It follows that scholars of computer-mediated behavior need methods for analyzing discourse, alongside

traditional social science methods such as experiments, interviews, surveys, and ethnographic observation.

This chapter describes an approach to researching online interactive behavior known as Computer-Mediated Discourse Analysis (CMDA). CMDA applies methods adapted from language-focused disciplines such as linguistics, communication, and rhetoric to the analysis of computer-mediated communication (Herring, 2001). It may be supplemented by surveys, interviews, ethnographic observation, or other methods; it may involve qualitative or quantitative analysis; but what defines CMDA at its core is the analysis of logs of verbal interaction (characters, words, utterances, messages, exchanges, threads, archives, etc.). In the broadest sense, any analysis of online behavior that is grounded in empirical, textual observations is computer-mediated discourse analysis.[3]

The specific approach to computer-mediated discourse analysis described here is informed by a linguistic perspective. That is, it views online behavior through the lens of language, and its interpretations are grounded in observations about language and language use. This perspective is reflected in the application of methodological paradigms that originated in the study of spoken and written language, e.g., conversation analysis, interactional sociolinguistics, pragmatics, text analysis, and critical discourse analysis. It also shapes the kinds of questions that are likely to get asked. Linguists are interested in language structure, meaning, and use, how these vary according to context, how they are learned, and how they change over time.

CMDA can be used to study micro-level linguistic phenomena such as online word-formation processes (Cherny, 1999), lexical choice (Ko, 1996; Yates, 1996), sentence structure (Herring, 1998), and language switching among bilingual speakers (Georgakopoulou, in press; Paolillo, 1996). At the same time, a language-focused approach can be used to address macro-level phenomena such as coherence (Herring, 1999a; Panyametheekul, 2001), community (Cherny, 1999), gender equity (Herring, 1993, 1996a, 1999b) and identity (Burkhalter, 1999), as expressed through discourse. Indeed, the potential—and power—of CMDA is that it enables questions of broad social and psychological significance, including notions that would otherwise be intractable to empirical analysis, to be investigated with fine-grained empirical rigor. The present chapter is intended as a practical contribution toward helping researchers realize this potential.

Because of its practical focus, this chapter will be most useful to readers who already have some study of computer-mediated communication in mind and who have given some thought to how they might approach their investigation. Readers who have made preliminary observations about a behavior (or behaviors) of interest in a specific online environment, and who have collected (or have access to) a relevant corpus of data, will be even better positioned to appreciate the methodological concerns addressed here. At the same time, the chapter is not intended as a step-by-step "how to" guide, but rather as an overview of how a CMDA researcher might conceptualize, design and interpret a research project involving identifying and counting discourse phenomena in a corpus of computer-mediated text.[4] For details regarding the implementation of specific analytic methods, readers are referred to the research studies cited in the references.

I begin by providing some historical background on CMDA and the kinds of research that have been carried out in the linguistic CMDA tradition, broadly construed. I then present a detailed overview of one version of the CMDA approach based on the "coding and counting" paradigm of classical content analysis, identifying a set of

conceptual skills necessary for carrying out a successful analysis. These skills are illustrated with reference to the problem of analyzing "virtual community" in two professional development sites on the Internet. In concluding, the limits of the coding and counting paradigm, and the CMDA approach as a whole, are identified and future directions are charted.

Background

The term "computer-mediated discourse analysis" was first coined in 1995 (see Herring, 2001), although research meeting the definitional criteria for CMDA has been carried out since the mid-1980s (in the linguistic sense: e.g., Murray, 1985, 1988; Severinson Eklundh, 1986), and arguably, as early as the 1970s (in the general sense: Hiltz & Turoff, 1978). Starting in the mid-1990s, and corresponding to the upsurge in computer-mediated communication (CMC) research that followed closely on the heels of the popularization of the Internet (Herring, 2002), an increasing number of researchers began focusing on online discourse as a way to understand the effects of the new medium. However, different researchers approached computer-mediated discourse with different questions, methods, and understandings, often working in isolation from one another—and in the case of researchers outside the United States, unaware that other researchers shared their interests. The present chapter attempts to systematize some of the goals, understandings, and procedures implicitly shared by this emerging cadre of researchers.

As background to the remainder of the chapter, it is useful to think of CMDA as applying to four domains or levels of language, ranging prototypically from smallest to largest linguistic unit of analysis: 1) structure, 2) meaning, 3) interaction, and 4) social behavior. Structural phenomena include the use of special typography or orthography, novel word formations, and sentence structure. At the meaning level are included the meanings of words, utterances (e.g., speech acts) and larger functional units (e.g., 'macrosegments', Herring, 1996b; cf. Longacre, 1992). The interactional level includes turn-taking, topic development, and other means of negotiating interactive exchanges. The social level includes linguistic expressions of play, conflict, power, and group membership over multiple exchanges. In addition, participation patterns (as measured by frequency and length of messages posted and responses received) in threads or other extended discourse samples constitute a fifth domain of CMDA analysis.

The kinds of understandings obtainable through a language-focused approach can be illustrated by summarizing briefly a few studies that focus on phenomena from each domain. Non-standard spelling and typography have been analyzed structurally in Internet Relay Chat as an example of creative play (Danet et al., 1997), on the French Minitel system as an illustration of the tension between efficiency and expressivity (Livia, in press), and in a social MUD as evidence of participants' "insider" status (Cherny, 1999). Studies that consider what online participants mean by what they say—for example, by classifying their utterances as speech acts—have discovered differences between educational and recreational uses of IRC, as well as differences associated with teacher/leader vs. other roles (Herring & Nix, 1997). Studies of interactional phenomena have identified system-imposed constraints on turn-taking (Herring, 1999a; Panyametheekul, 2001) and topic coherence (Herring & Nix, 1997; Lambiase, in press). One stream of socially-focused CMDA, research on group identity, has identified

discourse styles associated with participant age (Ravert, 2001), gender (Hall, 1996; Herring, 1993, 1996a, b, in press a), ethnicity (Paolillo, in press) and race (Burkhalter, 1999; Jacobs-Huey, in press), even in supposedly anonymous text-only CMC. Finally, participation patterns have been observed to vary according to the synchronicity of the medium (Condon & Cech, 2001, in press), and to reveal social influence and dominance in online groups (Herring, in press b; Herring et al., 1992; Hert, 1997; Rafaeli & Sudweeks, 1997). This brief survey is intended to provide a sense of the range and diversity of topics that have been researched thus far using CMDA. More detailed surveys of the findings of previous CMDA research can be found in Herring (2001, 2002).

The CMDA Approach

CMDA is best considered an approach, rather than a "theory" or a single "method". Although the linguistic variant described here is based on a loose set of theoretical premises (those of linguistic discourse analysis, plus a rejection of *a priori* technological determinism; see below), it is not a theory in that CMDA (as an abstract entity) makes no predictions about the nature of computer-mediated discourse. The findings of CMDA studies neither support nor falsify the premises of the approach, beyond confirming that it is useful or indicating that it is in need of further refinement. Rather, the CMDA approach allows diverse theories about discourse and computer-mediated communication to be entertained and tested. Moreover, although its overall methodological orientation can be characterized (see below), it is not a single method but rather a set of methods from which the researcher selects those best suited to her data and research questions. In short, CMDA as an approach to researching online behavior provides a methodological toolkit and a set of theoretical lenses through which to make observations and interpret the results of empirical analysis.

The theoretical assumptions underlying CMDA are those of linguistic discourse analysis, broadly construed. First, it is assumed that *discourse exhibits recurrent patterns*. Patterns in discourse may be produced consciously or unconsciously (Goffman, 1959); in the latter case, a speaker is not necessarily aware of what she is doing, and thus direct observation may produce more reliable generalizations than a self-report of her behavior. A basic goal of discourse analysis is to identify patterns in discourse that are demonstrably present, but that may not be immediately obvious to the casual observer or to the discourse participants themselves. Second, it is assumed that *discourse involves speaker choices*. These choices are not conditioned by purely linguistic considerations, but rather reflect cognitive (Chafe, 1994) and social (Sacks, 1984) factors. It follows from this assumption that discourse analysis can provide insight into non-linguistic, as well as linguistic, phenomena. To these two assumptions about discourse, CMDA adds a third assumption about online communication: *computer-mediated discourse may be, but is not inevitably, shaped by the technological features of computer-mediated communication systems*. It is a matter for empirical investigation in what ways, to what extent, and under what circumstances CMC technologies shape the communication that takes place through them (Herring, u.c.).

The basic methodological orientation of CMDA is language-focused content analysis. This may be purely qualitative—observations of discourse phenomena in a sample of text may be made, illustrated, and discussed—or quantitative—phenomena may

be coded and counted, and summaries of their relative frequencies produced. (It should be noted that quantitative CMDA comprises a qualitative component, e.g., in deciding what counts as an instance of a phenomenon to be coded and counted, especially when the phenomena of interest are semantic rather than syntactic (structural) in nature; see Bauer, 2000, and "analytical methods", below).

An example of the quantitative approach is Simeon Yates' (1996) comparison of a corpus of asynchronous computer conferences with spoken and written English corpora with respect to range of vocabulary, modality, and personal pronoun use. An example of the qualitative approach is Lori Kendall's (2002) ethnographic, participant-observer study of gendered behavior in a social MUD. An earlier ethnography of a social MUD carried out by Lynn Cherny (1999) applies both approaches, but to different phenomena: qualitative description of novel word creations (Ch. 3) and quantitative analysis of turn-taking patterns (Ch. 4). Alternatively, Herring (1996b) combines the two approaches: the same patterns of email message structure are identified by both qualitative and quantitative means.[5]

As with other forms of content analysis, the CMDA researcher must meet certain basic requirements in order to conduct a successful (i.e., valid, coherent, convincing) analysis. She must pose a research question that is in principle answerable. She must select methods that address the research question, and apply them to a sufficient and appropriate corpus of data. If a "coding and counting" approach is taken, she must operationalize the phenomena to be coded, create coding categories, and establish their reliability, e.g., by getting multiple raters to agree on how they should be applied to a sample of the data. If statistical methods of analysis are to be used, appropriate statistical tests must be identified and applied. Finally, the findings must be interpreted responsibly and in relation to the original research question. These requirements have been discussed extensively in the literature on the conduct of empirical research (see, e.g., Alford, 1998 for research in sociology; Bauer, 2000 for content analysis methods in communication); a basic familiarity with them is assumed here. Of interest in the present chapter is how to apply this general research schema to the particular constellation of issues and challenges associated with the study of computer-mediated behavior.

As an illustration of the CMDA approach, the following sections consider a currently popular research theme—that of "virtual community"—and how CMDA can be applied to determine empirically whether a group of people interacting online constitutes a community. In keeping with the focus of this volume on *learning*, the two online environments chosen for illustration have professional development as their reason for existence and both are associated with educational contexts: secondary science and mathematics education in the first case, and tertiary linguistics education and research in the second. To address the volume's focus on *system design*, the environments were selected to contrast in their technological affordances (one is a multimodal Web site, the other a text-based listserv); furthermore, one was intentionally designed with the goal of creating community, whereas the other was not. A comparison of these two environments can shed light on how the technological and social properties of CMC systems relate to the phenomenon of virtual community.

Analyzing "Virtual Community"

Since it was first articulated in print (Rheingold, 1993), the concept of "virtual community" has become increasingly fashionable in Internet research (e.g., Baym, 1995a; Cherny, 1999; Werry & Mowbray, 2001), although it has also been criticized (Fernback & Thompson, 1995; Jones, 1995a; see also Kling & Courtright, this volume). The criticisms include a pragmatic concern that the term has been overextended to the point of becoming meaningless—for some writers, it seems that any online group automatically becomes a "community"—and a philosophical skepticism that virtual community can exist at all, given the fluid membership, reduced social accountability, and lack of shared geographical space that characterize most groups on the Internet (e.g., McLaughlin et al., 1995). For the purposes of the present discussion, we assume that virtual community is possible, but that not all online groups constitute virtual communities. The task of the researcher then becomes to determine the properties of virtual communities, and to assess the extent to which they are (or are not) realized by specific online groups.

Two Learning Environments

Two online professional development environments will serve as examples to ground our discussion of how CMDA can be applied to investigate virtual community. Professional development environments are online learning environments in which people participate voluntarily and intermittently—i.e., for the purpose of acquiring information and skills to advance professionally—rather than in formal courses with students, instructors, and syllabi, as is the case for distance education. In successful cases, participation in such environments is continuous and self-sustaining, unlike course-based CMC, which is task-focused and temporally bounded. An example of a genre of professional development environment that dates back to the early days of computer networking is listserv discussion groups for professionals in academic disciplines (e.g., Hert, 1997; Korenman & Wyatt, 1996). A more recent example is the growing genre of professional development Web sites that combine discussion forums with access to documents and other online resources (e.g., Renninger, this volume).

      The environments selected as illustrations for this chapter represent these two types. The first, the Linguist List, was founded in November 1990 by a husband and wife team of academic linguists as a means for disseminating information and engaging in public discussion about issues of interest to professional (and aspiring professional) linguists; it has been in continuous existence since 1990. Originally a text-only, by-subscription list that made archived messages available only to subscribers, in 1994 it established a Web site and posted the discussion archives there, making them widely publicly accessible.[6] For further description and analysis of the Linguist List, see Herring (1992, 1996b).

      The second environment, the Inquiry Learning Forum (ILF), was opened to registered members in March 2000. It was designed with National Science Foundation support by a team of faculty and graduate students in the School of Education at Indiana University, with the explicit goal of fostering online community among secondary math and science in-service and pre-service teachers interested in the inquiry learning approach (National Research Council, 2000). Members must go to the ILF Web site to post

messages and access the other resources there (which include videos of teachers using inquiry methods in their classrooms); past messages remain on the site alongside current messages. For further description and analysis of the ILF, see Barab, MaKinster, & Scheckler (this volume) and Herring, Martinson & Scheckler (2002).

These environments are plausible candidates for virtual community status in several respects. First, both bring together people who arguably already constitute real-world professional communities: academic linguists and secondary math and science educators. Second, their online participation is centered around a shared professional focus, as in Wenger's (1998) "communities of practice." Third, the Linguist List is active and long-lived, which some might take as prima facie evidence that it has achieved online community status. In contrast, the ILF has struggled to establish and maintain an active level of participation, but might be considered to have a prima facie claim to community status on the grounds that it was explicitly designed to support community (Barab, MaKinster, Moore, Cunningham, & The ILF Design Team, in press). For these reasons, it is germane to ask: To what extent does participation in these two environments in fact constitute "community" (as opposed to being simply "people interacting online")?

The following sections describe how a researcher making use of CMDA might go about addressing this question. Five conceptual skills involved in the research process are highlighted and discussed, first, with reference to CMDA in general, second, with reference to virtual communities, and last, with reference to the two professional development sites. The order of presentation of the five skills is roughly sequential (i.e., a researcher generally starts with the first, and progresses to the last), although the research process—in CMDA, no less than in other scientific disciplines—is frequently iterative, involving many feedback loops (Harwood et al., 2001). However, it is important to stress that what follows is not intended as an analysis in and of itself; to answer the question of what constitutes online community definitively would take us well beyond the scope of the present chapter.

Research Questions

To carry out an investigation by means of CMDA, it is first necessary to have a research question, a problem to which the analyst desires to find a solution. Typically, the research question is based on prior observation—the researcher may have noticed some online behavior or behaviors and may have formed a preliminary hypothesis concerning them. Articulating a research question is a first step towards testing the hypothesis.

A good CMDA research question has four characteristics:

1) It is empirically answerable from the available data;
2) it is non-trivial;
3) it is motivated by a hypothesis; and
4) it is open-ended.

Each of these characteristics is discussed below.

A CMDA research question should ideally ask about *empirically-observable phenomena*, or phenomena that can be operationalized empirically, as opposed to purely subjective or evaluative ones. A question about the nature and frequency of joking in an online forum, for example, can be addressed empirically more readily than a question

about whether the participants are having fun. Further, the question should be *answerable from the data selected for analysis*. For example, if only computer-mediated data are to be examined, the question should not ask whether CMC is better or worse than face-to-face communication along some dimension of comparison, since the CMC data can not tell us anything directly about face-to-face communication. Equally important in CMDA, the question should be *answerable on the basis of textual evidence*. Text is direct evidence of behavior, but it can only be indirect evidence of what people know, feel, or think. If it is important that the researcher try to understand participants' internal conscious or unconscious states, CMDA should be supplemented with other methods of analysis such as interviews or psychological experiments.

A good research question should be *non-trivial*; that is, the answer should be *of some ostensible interest* to at least a portion of the larger research community, and *not already known* in advance. Additionally, the research question should not be worded so as to presuppose an answer; that is, the answer should not appear to be a foregone conclusion.

At the same time, a research question *motivated by a hypothesis*—even if it is no more than an informal hunch—is more interesting and more interpretable than one that is not. Note that it is not necessary to posit a hypothesis that the researcher expects will be confirmed by the results of the analysis, although the hypothesis should be prima facie plausible. In some cases, a researcher may advance a popular hypothesis that she suspects is incorrect, in order to disprove it. For example, she might postulate that participant gender is invisible in CMC (a commonly held view in the early 1990s, based on the paucity of social status cues in text-only CMC), suspecting that such is not the case in her data.[7] The empirical results, if negative, are all the more illuminating for running counter to the prevailing wisdom.

Ideally, whether the researcher's hypothesis is supported or not, the results of the study should contribute new knowledge. Phrasing the question as an *open-ended* question (what, why, when, where, who, how) leaves the door open to unexpected findings to a greater extent than closed (yes/no) questions, generally speaking. One caveat is that unexpected answers to yes/no questions can be informative, as noted above, when the hypothesis underlying the question is favored by popular opinion or common sense, but receives no empirical support. Similarly, positive support for an unobvious hypothesis can also cause us to understand the world in new ways. However, support for obvious hypotheses does not advance knowledge, nor does lack of support for unobvious hypotheses. In contrast, a systematic study will always reveal something new in response to a well-crafted "what", "why", or "how" question.

What kinds of questions about virtual community can be researched from a CMDA perspective? Although all are legitimate foci of intellectual curiosity, the researcher is setting herself up for difficulty if she asks questions such as: i) "Does virtual community exist?" ii) "Is virtual community a good thing?" iii) "Does membership in virtual communities satisfy needs previously satisfied only in face-to-face communities?" or iv) "Do people interact regularly in groups online?" Note, first of all, that these are closed questions, to which the answer can only be "yes" or "no". In addition, the first is effectively biased towards an affirmative answer, in that exhaustive evidence would be required in order to answer it negatively. The second question both presupposes the existence of virtual community (a problem if virtual community hasn't already been empirically demonstrated) and asks a subjective, evaluative question about it; "goodness" is difficult to

measure empirically. The third question involves a comparison; it can only be answered if empirical evidence (gathered by comparable means) is available from both "virtual communities" (presupposed to exist) and face-to-face communities. Finally, the fourth question, although neutrally worded and answerable, is trivial—the answer is obvious to anyone who has spent any time on the Internet.

The following, in contrast, are examples of open-ended questions that can usefully be addressed using CMDA: a) "What are the discourse characteristics of a virtual community?" (b) "What causes an online group to become a community?" c) "What causes a virtual community to die?" d) "How do virtual communities differ from face-to-face communities?"[8] e) "What happens to face-to-face communities when they go online?" and f) "In what ways do communities constituted exclusively online differ from online communities that also meet face-to-face?" However, these questions are not all equally easy to answer; their answerability depends on the data available for investigation. Thus, for example, a)-d) and f) require an independent determination of virtual community, e.g., in terms of participants' perceptions; b), c), and e) require longitudinal data; and d) and e) require face-to-face data (see discussion of "data" below).

In addition, particular data samples will generally exhibit characteristics that invite more specific questions to be asked about them. The question raised in the previous section—"[t]o what extent does participation in these two environments constitute 'community' (as opposed to being simply 'people interacting online')?"—is a straightforward application of question (a) to the Linguist List and the ILF data samples. But these samples, by their nature, also give rise to questions about virtual community and professional development (e.g., "What is the nature of virtual community in professional development environments, and how does it differ from virtual community in structured learning environments / unstructured social environments / etc.?"). Furthermore, the two environments contrast according to a number of technological and social dimensions, as summarized in Table 1.[9] Additional questions can be asked to focus on the contributing effects of a particular dimension to online behavior (e.g., "Is a multimodal environment more conducive to virtual community than a text-only environment?"; or "How does the self-presentation of the group 'owners' (e.g., as peers or as experts) affect the likelihood that a group will develop community characteristics?").

The comparison of the two groups in Table 1 suggests *too* many possible questions about the variables that condition virtual community, in fact. Ideally, two data samples that are compared should differ according to only one dimension, such that if differences in behavior are found between the samples, they can plausibly be attributed to that dimension of variation. If, however, it turns out that either the Linguist List or the ILF exhibits more "community" behaviors than the other, to what should the difference be attributed: (multi)modality? ease of posting messages? ease of access to the group's history? availability of face-to-face interaction? the intentions/behavior of the group's founders? etc. Causal indeterminacy is a common problem in research that analyzes naturally occurring behavior.[10] The experimental research paradigm controls for this by holding all variables constant except for the variable that is hypothesized to condition the experimental result. For examples of experimental research that make use of CMDA methods, see Condon & Cech (1996a, 1996b, 2001).

Table 1. Dimensions of contrast between the Linguist List and the ILF

Linguist List	**ILF**
Text-only	Multimodal (text + video + limited audio and graphics)
Messages come to subscriber ("push" technology)[11]	Member must go to site to post messages ("pull" technology)
Archives stored separately	Past messages appear alongside current ones
Public (by subscription)	Semi-public (by registration; password required; limited membership)
Pre-existing face-to-face "community" (meets at annual professional meeting)	Loosely defined pre-existing "community" (most members have never met face-to-face)
Relatively homogeneous population of users (academic linguists at universities) with similar access opportunities	Heterogeneous population of users (pre-service teachers; in-service teachers; ILF researchers) with differential access
Founders' goals were specific, limited in scope (i.e., information exchange & discussion)	Creators' goals were broad, ambitious (i.e., create intentional community; foster inquiry learning)
Moderators present themselves as peers, "facilitators" (but exercise behind-the-scene control over postings)	ILF development team members have higher status (but post messages themselves, and do not control postings)
Discussion is on topics selected by participants	Discussion is often focused around artifacts (video clips; instructional technology; lesson plans, etc.)

Data Selection

In CMDA, as in other empirical social science approaches, a data sample must be selected that is appropriate to the study. By "appropriate" is meant that the sample should be of a nature and size to answer the research question(s); if the research question involves a comparison, more than one sample may be required. Each of these considerations is discussed below. For the purposes of this discussion, it is assumed that the data of interest are produced naturally (i.e., by online discourse participants for their own purposes), and logged or culled from online archives by the researcher, rather than elicited experimentally.

It is often impossible to examine all the phenomena of relevance to a particular research question; this is especially true in CMDA, for which a vast amount of textual data is available in the form of online interactions. (Even in groups with relatively low participation, such as the ILF in its first year, the total amount of text quickly adds up to more than can easily be analyzed by a human coder using micro-linguistic methods.) For this reason, the researcher must usually select a sample from the totality of the available data. In CMDA, this is rarely done randomly, since random sampling sacrifices context, and context is important in interpreting discourse analysis results. Rather, data samples tend to be motivated (e.g., selected according to theme, time, phenomenon, individual or group), or samples of convenience (i.e., what the researcher happens to have access to at the time). Some advantages and disadvantages of these various sampling techniques are summarized in Table 2.

Table 2. CMDA data sampling techniques

	Advantages	Disadvantages
Random (e.g., each message selected or not by a coin toss)	representativeness; generalizability	loss of context & coherence; requires complete data set to draw from
By theme (e.g., all messages in a particular thread)	topical coherence; a data set free of extraneous messages	excludes other activities that occur at the same time
By time (e.g., all messages in a particular day/week/month)	rich in context; necessary for longitudinal analysis	may truncate interactions, and/or result in very large samples
By phenomenon (e.g., only instances of joking; conflict negotiation)	enables in-depth analysis of the phenomenon (useful when phenomenon is rare)	loss of context; no conclusions possible re: distribution
By individual or group (all messages posted by an individual or members of a demographic group, e.g., women, students)	enables focus on individual or group (useful for comparing across individuals or groups)	loss of context (especially temporal sequence relations); no conclusions possible re: interaction
Convenience (whatever data are available to hand)	convenience	unsystematic; sample may not be best suited to the purposes of the study

Of the techniques in Table 2, temporal sampling preserves the richest context. If a long enough continuous time period is captured, the sample will most likely include coherent threads, thereby incorporating the advantages of thematic sampling as well. Analogously, a thematic sample is typically organized by time, enabling some longitudinal

observations to be made. Because of their multiple advantages, these two sample types are favored in CMDA research. In addition, it is possible to break a sample of any type down by individual or group, thereby achieving additional focus while avoiding the disadvantages of individual or group sampling. (For example, an extended thread was isolated for analysis from the Linguist List, then broken down by gender of participants, in Herring, 1992, 1996b).

The richest possible context is required for the purposes of analyzing virtual community, as are data that can show change over time, if questions about the inception, evolution, and demise of virtual communities are to be addressed. The sample should include, as much as is possible, the typical activities carried out on the site. These considerations suggest intermittent time-based sampling (e.g., several weeks at a time at intervals throughout a year) as particularly appropriate.[12] Ideally, in any analysis of virtual community, textual analysis would be supplemented by ongoing participant observation.[13]

The ILF environment imposes some limitations on sampling, as well as suggesting alternative sampling possibilities. Discussions take place in different parts of the ILF site, making it difficult to capture a representative overall time-based sample; rather, samples must be collected from individual "rooms" and collated, if a single sample is required. Moreover, discussions in the "classroom" portion of the ILF site are organized around videos of teachers using inquiry methods in their classrooms, with one discussion forum attached to each video (Herring et al., 2002). This configuration suggests new categories of data sampling: by room, and by artifact (in this case, video). A sampling technique based on units of interaction determined by the site design (and/or by participants' actual usage) has the advantage of allowing discourse patterns to emerge that are internally coherent to such units, whereas if data are combined across units, those patterns might be less apparent.

How much data is required to conduct a successful CMDA study? There is no simple answer to this question. The data should be sufficient to address the research question, such that tests of statistical significance could meaningfully be conducted on the key findings (regardless of whether or not the researcher actually conducts such tests). What counts as a sufficient amount of data will depend, therefore, on the frequency of occurrence of the analytical phenomenon in the data sample, the number of coding categories employed to describe the phenomenon, and the number of external factors that are allowed to vary (e.g., modality; topic of discussion; participant gender). Two general rules of thumb are 1) the more infrequent the phenomenon in the data, the larger the sample should be, and 2) the more variables considered in the analysis, the larger the sample should be. This is so that 1) enough instances of the phenomenon are available to analyze, and 2) when the sample is broken down into sub-samples for purposes of comparison, there are still enough instances in each category to allow for statistical testing.[14] Since it is often difficult to know all of this in advance, a recommended practice is to start with a pilot study based on a small amount of data, and expand the sample size as necessary in a larger study, according to the tendencies revealed in the pilot study.

A related issue concerns the number of samples required for purposes of comparative analysis. Above we noted that some CMDA research questions presuppose a comparison with face-to-face discourse. While it may be legitimate to draw a comparison with previous research on face-to-face communication in interpreting one's results (see "interpretation" below), no key results should be founded on such a comparison, unless the

researcher can assure that the face-to-face study was carried out using comparable methods (e.g., because it was conducted by the researcher herself, or because the same methods that were applied in the face-to-face study were applied to the computer-mediated data). Otherwise, a comparable face-to-face sample is normally required. What the researcher hopes to find are cases in which the same people are communicating about the same topics, for the same purposes, both face-to-face and via CMC. Unfortunately, this situation rarely occurs naturally. Left to their own devices, people tend to use different modalities for different communicative purposes; moreover, CMC enables certain behaviors that would be difficult or impossible offline,[15] and vice versa. Data collected in experimental settings are superior to naturally-occurring data for the purposes of comparing CMC with face-to-face (and traditional written) communication (see, e.g., Condon & Cech, 1996a, 1996b, 2001). However, since evidence of community is highly unlikely to surface in laboratory settings, given that experimental subjects typically have no past (or anticipated future) interaction (Walther, 1996), empirical comparison of face-to-face and online community is difficult. This may be one question for which interpretive, rather than strictly empirical, answers will have to suffice for the present time (cf. Etzioni, 1999).

Multiple CMC samples (or sub-samples) may also be required in order to carry out a single study, depending on the research question. These are usually easier to collect, but care should be taken to hold constant as many dimensions of variation as possible, to maximize the interpretability of the results. Our two professional development samples in fact vary according to too many dimensions to enable straightforward comparison, as noted above. A better example of contrasting samples is Paolillo's (in press) comparison of a(n asynchronous) Usenet newsgroup and a (synchronous) IRC channel frequented by the same participant demographic group (and to some extent, the same individuals): expatriate South Asians. When differences are found in language choice in the two samples, they can plausibly be attributed to differences in synchronicity between the two CMC modes.

Dividing a larger sample into sub-samples by demographic group, topic, or other category is another means to insure that the sub-samples share all but one feature. Applying this principle to research on virtual community, we might, for example, compare the behaviors of individuals within a single group who are known to interact face-to-face with other group members, with those individuals who do not, to test the hypothesis that face-to-face contact enhances involvement in online community (cf. Diani, 2000). Or we might consider participant behavior by role or status in relation to hypothesized community behaviors. In the case of the Linguist List, the behavior of professors might be compared with that of students, or U.S. linguists with non-U.S. linguists; in the ILF, pre-service teachers might be compared with in-service teachers, and teachers with researchers, to determine if higher status groups are more invested in the "community" than lower status groups.[16]
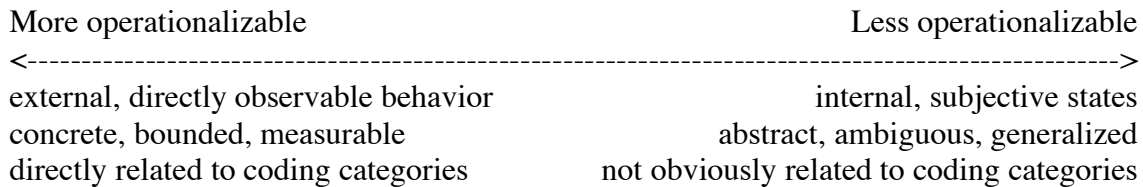
Operationalization of Key Concepts

The coding and counting approach to CMDA research described in this chapter requires that key concepts be operationalizable (and operationalized) in empirically measurable terms. This entails defining the concepts unambiguously, such that another researcher, examining the same data, could in principle reproduce the identification of a given token as an exemplar of the concept.[17] Equally or more important, it is necessary to define a concept

in concrete, textual terms in order to be able to code it consistently. In the case of highly abstract concepts, this necessarily entails a reduction (and a risk of distortion) of the concept; content analysis is sometimes criticized on these grounds (cf. Bauer, 2000). At the same time, it is the requirement of operationalization, more than any other single requirement, that lends CMDA its rigor and makes it a useful tool for getting an empirical grasp on otherwise slippery or intractable concepts.

Concepts vary in the degree to which they are inherently operationalizable. This can be represented as a continuum, as in Figure 1. In a previous section, it was suggested that a researcher should avoid asking questions about concepts that are too far towards the subjective, abstract end of the continuum. In fact, such questions are often the most interesting to ask, but in order to address them quantitatively using CMDA, they must be defined in terms of textual phenomena that can be directly observed, coded, and counted. Thus, for example, concepts of widespread interest in CMC research such as affect, democracy, depth (of discussion), empowerment learning, trust, etc. can be operationalized by identifying discourse behaviors (plausibly) characteristic of each phenomenon and then articulating interpretive links between those behaviors and the larger concepts. (We will see how this might be done for the concept of virtual community below.) Alternatively, it might be necessary to supplement CMDA with other methods in order to make a meaningful demonstration that the evidence addresses the concept. For example, it is unlikely that CMC evidence alone could make a definitive case for changes in offline states of affairs; such a demonstration would normally require offline evidence, observational or self-reported.

Figure 1. Continuum of operationalizability

More operationalizable   Less operationalizable
<-->
external, directly observable behavior   internal, subjective states
concrete, bounded, measurable   abstract, ambiguous, generalized
directly related to coding categories   not obviously related to coding categories

"Community" is an inherently abstract concept. It also has a subjective component, especially when it is applied to online contexts, where it is always, in some sense, a metaphorical extension of the literal meaning of community as "grounded in a shared physical space" (cf. Jones, 1995a). Accordingly, definitions of community (and virtual community) abound, although Wellman's (2001) tripartite characterization of community as providing "sociability, support, and identity" constitutes a useful point of departure. More specifically, six sets of criteria can be identified from the literature on virtual community (e.g., Haythornthwaite et al., 2000; Jones, 1995a, 1995b; Reid, 1991, 1994, 1998; Riel, this volume):

   1) active, self-sustaining participation; a core of regular participants
   2) shared history, purpose, culture, norms and values
   3) solidarity, support, reciprocity
   4) criticism, conflict, means of conflict resolution
   5) self-awareness of group as an entity distinct from other groups
   6) emergence of roles, hierarchy, governance, rituals

Criteria 1) and 4) relate to "sociability"; criteria 3) and 6) (loosely) to "support", and criteria 2) and 5) to "identity."[18]

These six criteria suggest concrete ways in which the notion of "virtual community" might be broken down into component behaviors that can be objectively assessed.

1) *Participation* can be measured over time, and *core participants* identified on the basis of frequency of posting and rate of response received to messages posted (Herring, in press b), or via text-based social network analysis (Paolillo, 2001; cf. Koku & Wellman, this volume).

2) Shared *history* can be assessed through the availability and use of archives (Millen, 2000). *Culture* is indexed through the use of group-specific abbreviations, jargon, and language routines (Baym, 1995a; Cherny, 1999; Jacobs-Huey, in press; Kendall, 1996), as well as through choice of language, register, and dialect (Georgakopoulou, in press; Paolillo, 1996). *Norms and values* are revealed through an examination of netiquette statements (Herring, 1996a), FAQs (Voth, 1999) and verbal reactions to violations of appropriate conduct (McLaughlin et al., 1995; Weber, in press).

3) *Solidarity* can be measured through the use of verbal humor (Baym, 1995b); *support* through speech act analysis focusing, e.g., on acts of positive politeness (Herring, 1994); and *reciprocity* through analysis of turn initiation and response (Rafaeli & Sudweeks, 1997).

4) *Criticism* and *conflict* can be analyzed through speech acts violating positive politeness (Herring, 1994). *Conflict resolution* might usefully be considered as an interactive sequence of acts (cf. Condon & Cech, 1996b on decision-making sequences); it also lends itself to ethnographic analysis (e.g., Cherny, 1999).

5) A group's *self-awareness* can be manifested in its members' references to the group as a group, and in 'us vs. them' language, particularly in statements to the effect, "We do things this way here" (implying an awareness that they might be done differently elsewhere; Weber, in press). (See also "norms" above.)

6) Evidence of *roles* and *hierarchy* can be adduced through participation patterns (see "participation" above) and speech act analysis (e.g., Herring & Nix, 1997, which considers the different acts performed by group leaders and non-leaders). The study of *governance* and *ritual* would appear to require an ethnographic approach in which a group's practices are observed over time and described in terms of their meanings to participants (Cherny, 1999; Jacobson, 1996; Kolko & Reid, 1998). Note, however, that the reification of cultural practices in the form of governance and ritual appears to represent a relatively advanced stage of community (see, e.g., Dibbell's 1993 account of how this happened in LambdaMOO); thus it probably should not be taken as part of the basic definition of virtual community.

Some of the above features are more useful than others as potential indicators of virtual community on the Linguist List and the ILF. Certain features occur rarely or not at all in either group: language routines, code switching, humor, and governance and ritual. Their relative absence is due to a variety of circumstances, for example the professional (serious) focus of the groups, and the fact that their members are proficient in written English.[19] Other features occur only or nearly exclusively on the Linguist List, e.g., criticism, conflict, and netiquette statements.[20] Conversely, such features as participation patterns, reciprocity, indicators of group self-awareness, and evidence of roles and

hierarchy are evident in both and might usefully be assessed as community indicators for these environments.

Analytical Methods

Analytical methods in CMDA are drawn from discourse analysis and other language-related paradigms, adapted to address the properties of computer-mediated communication. In principle, nearly any language-related method could be so adapted; in practice, this chapter focuses on methods of linguistic discourse analysis, these being the methods with which the author is most familiar. These include approaches traditionally used to analyze written text and spoken conversation, approaches to discourse as social interaction, and critical (socio-political) approaches.

Given that we have already identified content analysis as the basic methodological apparatus of CMDA, the question might arise as to what the more specific linguistic approaches add to the research endeavor. In fact, it is possible to conduct a perfectly responsible CMDA analysis without drawing on any more specific paradigm than language-focused content analysis. For example, one could let the phenomenon of interest emerge out of a sample of computer-mediated data and devise coding categories on the basis of the observed phenomenon, as in the grounded theory approach (Glaser & Strauss, 1967). This approach is especially well suited to analyzing new and as yet relatively undescribed forms of CMC, in that it allows the researcher to remain open to the possibility of discovering novel phenomena, rather than making the assumption in advance that certain categories of phenomena will be found.

However, grounded theory is less useful for evaluating specific research hypotheses, or for making systematic comparisons across data samples. For these purposes, the CMDA researcher can profit from the structure, experience, and understandings available through specific discourse analysis paradigms. Such paradigms define issues of theoretical interest, a set of discourse phenomena about which much may already be known in other modalities and contexts, and discovery procedures for revealing the patterns and constraints that characterize the phenomena. Table 3 summarizes this information for five discourse analysis paradigms commonly invoked in CMDA research.

However, while it is useful to be cognizant of these research paradigms as part of the CMDA toolkit, and to draw on them as appropriate, most CMDA research does not take as its point of departure a paradigm, but rather observations about online behavior as manifested through discourse. That is, rather than starting off with the intention of using conversation analysis (for example) to investigate some aspect of CMC and then selecting a behavior to focus on, a researcher is more likely to become interested in studying patterns of message exchange (for example), and then select conversation analysis as a useful methodological tool. In this sense, the approach is inductive—the phenomena of interest are primary—rather than deductive, or theory-driven. This orientation is reflected in Table 4, in which essentially the same CMDA issues and methods are re-organized around the four domains of language (plus participation) identified at the beginning of this chapter. Each domain includes sub-sets of linguistic phenomena, listed in the second column of Table 4.

Table 3. Five discourse analysis paradigms

	Issues	Phenomena	Procedures
Text Analysis (cf. Longacre, 1996)	classification, description, "texture" of texts	genres, schematic organization, reference, salience, cohesion, etc.	identification of structural regularities within and across texts
Conversation Analysis (cf. Psathas, 1995)	interaction as a jointly negotiated accomplishment	turn-taking, sequences, topic development, etc.	close analysis of the mechanics of interaction; unit is the turn
Pragmatics (cf. Levinson, 1983)	language as an activity—"doing things" with words	speech acts, relevance, politeness, etc.	interpretation of speakers' intentions from discourse evidence
Interactional Sociolinguistics (cf. Gumperz, 1982; Tannen, 1993)	role of culture in shaping and interpreting interaction	verbal genres, discourse styles, (mis)communication, framing, etc.	analysis of the socio-cultural meanings indexed through interaction
Critical Discourse Analysis (cf. Fairclough, 1992)	discourse as a site in which power and meaning are contested and negotiated	transitivity, presupposition, intertextuality, conversational control, etc.	interpretation of meaning and structure in relation to ideology, power dynamics

Table 4. Four domains of language

	Phenomena	Issues	Methods
Structure	typography, orthography, morphology, syntax, discourse schemata	genre characteristics, orality, efficiency, expressivity, complexity	Structural/Descriptive Linguistics, Text Analysis
Meaning	meaning of words, utterances (speech acts), macrosegments	what the speaker intends, what is accomplished through language	Semantics, Pragmatics
Interaction	turns, sequences, exchanges, threads	interactivity, timing, coherence, interaction as co-constructed, topic development	Conversation Analysis, Ethnomethodology
Social behavior	linguistic expressions of status, conflict, negotiation, face-management, play; discourse styles, etc.	social dynamics, power, influence, identity	Interactional Sociolinguistics, Critical Discourse Analysis

Participation, while not a level of linguistic analysis per se, constitutes a fifth domain, in which the phenomena of interest are number of messages and responses and message and thread length. Such numbers can be interpreted to address social issues such as power, influence, engagement, roles, and hierarchy. Participation is not associated with a particular set of discourse analysis methods, but rather with descriptive statistics (i.e., the phenomena are simply counted).

Bauer (2000) draws a useful distinction in content analysis between "syntactic" (structural) and "semantic" phenomena. The former are invariant in form, or their members comprise a limited set of variants that can be formally identified. Examples of structural CMC phenomena include emoticons, abbreviations, lexical items (such as personal pronouns), word formatives (such as *cyber-*), syntactic patterns (such as passive voice), and quoting (when marked by a formal signal, such as quotation marks or an angle bracket > at the beginning of a line of text). Such phenomena are objectively identifiable; they can be coded and counted more or less automatically, on the basis of a predefined set of structural features. Obviously, these are advantages if the researcher wishes to conduct computer-assisted data analysis.

Semantic coding categories, in contrast, hold the meaning or function constant, but vary (sometimes endlessly) in form. Examples of semantic CMC phenomena include speech acts and most social phenomena such as conflict and politeness.[21] Coding such phenomena necessarily involves an interpretive, subjective component; in most cases it can only be carried out by human coders. Despite the greater challenges they pose for empirical investigation, semantic phenomena are often the most interesting to study.

Empirical rigor can be maintained if the researcher operationalizes and defines each coding category in explicit terms and applies the codes consistently to the data. To insure consistency of coding, inter-rater reliability measurements can be taken in CMDA, as in other forms of content analysis. This is especially advisable when the coding incorporates a subjective component.

The structural language phenomena in Table 4 are generally "structural" (or "syntactic") in Bauer's sense. Interactional phenomena such as threading (based on subject line) can also be identified on structural grounds. To the extent that key words identify social phenomena, the frequency of those words can be counted, making structural methods appropriate to some social questions as well. Word and message counts are purely structural. In contrast, meaning, most social phenomena, and any interactional phenomena that require interpretation are "semantic" in Bauer's sense. One practical consequence of the greater ease with which structural phenomena can be automated is that analyses of such phenomena can be carried out on large samples of data. Conversely, semantic analyses, because they must be done "by hand," effectively limit the amount of data that can be analyzed.[22]

In the discussion of "operationalization" above, various discourse behaviors were identified as possible indicators of virtual community. These represent both structural and semantic phenomena, and span all five domains of CMDA. Table 5 summarizes these behaviors.

Table 5. Discourse behaviors hypothesized to indicate virtual community

structure	jargon, references to group, in-group/out-group language
meaning	exchange of knowledge, negotiation of meaning (speech acts)
interaction	reciprocity, extended (in-depth) threads, core participants
social behavior	solidarity, conflict management, norms of appropriateness
participation	frequent, regular, self-sustaining activity over time

In an actual CMDA analysis of the evidence for virtual community in the Linguist List and the ILF, one or more behaviors would be selected from Table 5 and explicit coding categories devised for each. For example, in-group/out-group language might be operationalized structurally as the uses of first-person plural pronouns ("we", "us", etc.) in contrast to third-person plural pronouns ("they", "them", etc.); reciprocity might be operationalized interactionally as "response to previous message" or "response to previous message exchange" (cf. Rafaeli & Sudweeks, 1997); and solidarity might be operationalized in social terms as the occurrence of humorous utterances (which would, in turn, need to be explicitly defined). An investigation that attempted to address *all* of the

behaviors in Table 5 would probably not be feasible, since each behavior would need to be coded whenever it applies in a sufficiently large enough sample to achieve meaningful results for each demographic, temporal, topical, or other sub-division of the data that is being considered, for each of the two groups. Unless many of the features were coded automatically, the coding involved would be excessively time-consuming, and the results too numerous to present and discuss in an article-length work (although such a project might be appropriate in scope for a doctoral dissertation). In light of these constraints—and since in any event few studies are able to analyze *all* the possible evidence pertinent to a given research question—a researcher will normally select those features to code that she believes will produce the most valid and convincing results in relation to the research question, which in this case concerns the presence or absence of virtual community.

Although space and scope considerations prevent us from undertaking a full-fledged analysis of the hypothesized community behaviors in the Linguist List and the ILF in this chapter, a superficial consideration of the behaviors in Table 5 nonetheless reveals some differences between the two groups. The Linguist List has an explicit set of norms and guidelines for appropriate posting behaviors that are periodically posted to the list; such norms, if they exist on the ILF at all, are implicit. The Linguist List is characterized by regular conflict episodes, some of which are resolved behind the scenes by the moderators (see, e.g., Herring et al., 1995). Indeed, conflict was a feature of the Linguist List from the outset (Herring, 1992). In contrast, the ILF has virtually no conflict episodes. Perhaps most significantly, the Linguist List is active and self-sustaining; it grew rapidly from about 500 to 4000 subscribers in the first year, doubling to 8000 after a few years; today, at over 12,000 subscribers, message volume is so great as to overwhelm some subscribers, even when messages are consolidated and distributed as daily digests. In contrast, the ILF has had to work hard to recruit members—as of January 2002, the number was around 1000, most of them pre-service teachers who were required to subscribe as part of their course work at Indiana University—and most members do not post. If they do, they do not return to the site subsequently, and few exchanges turn into extended threads.

There are also similarities. Both sites make use of professional jargon; both reference themselves as an in-group in relation to an out-group (non-linguists; students); both exchange knowledge[23] (although more of this takes place on the Linguist List than on the ILF); and both make limited use of expressions of solidarity. In a quantitative study, these observations would be supported with numerical evidence of frequency distributions for each behavior, compared across the two sites. How might we interpret such evidence in relation to the question of whether the two environments are virtual communities?

Issues of Interpretation

Responsible interpretation of research findings is necessary to insure the validity of any study. Skillful interpretation, moreover, makes the difference between a competent investigation and an insightful one. Interpretation is thus both a craft and an art. Interpretation of the results of CMDA should ideally take into account medium and situational variables, and take place on three levels: close to the data, close to the research question, and (optionally) beyond the research question.

Medium and situational variables are dimensions according to which computer-mediated data can vary and which potentially condition significant variation in online behavior. An example of a medium variable is synchronicity; an example of a situational variable is participant demographics (for a longer list of variables of each type, see Herring, u.c.). Such variables often enter into decisions about data selection early in the research process and can function as explicit dimensions of contrast within a study—for example, a synchronous sample may be compared with an asynchronous sample; native English speakers may be compared with non-native English speakers, males with females, teachers with students, and so forth. These same dimensions are also relevant in interpreting analytical results, even in studies with relatively homogeneous data sets.

The issue is one of generalizability of the research findings: for what kinds of CMC—beyond the specific sample(s) analyzed—might the findings hold true? Strictly speaking, every sample is unique, and thus all generalization should be undertaken with caution. At the same time, results that do not generalize beyond the sample in the study are less valuable and interesting than those that do, a consideration that argues against excessive conservatism in interpretation. Advancing explanations that take into account medium and situational variables is one way to balance these competing requirements.

Another strategy for balancing caution with generalization is to interpret the research findings at multiple levels. *Interpretation close to the data* involves summarizing and synthesizing the results obtained by applying the analytical methods to the data. At this most conservative level of interpretation, patterns of results should be identified. *Interpretation close to the research question* requires the researcher to revisit the research questions raised at the outset of the study and indicate explicitly how the results answer the questions. Some creative reasoning may be required here; for example, the steps necessary to reason from the larger concepts in the research question to the specific, operationalizable features of the text may need to be reversed. At this level of interpretation, the researcher should also point out which results are expected and which are unexpected, and propose explanations for the unexpected results. The third and *broadest level of interpretation* calls upon the researcher to extrapolate from the findings of the study to their theoretical, methodological, and/or practical (e.g., design) implications. This level is necessarily the most speculative, and is not strictly speaking required to complete a study. However, broader interpretation helps others to appreciate the significance of the analysis, and can suggest productive avenues for further research.

Because interpretation is a creative intellectual act and because there can be more than one possible (broad) interpretation for any given analytical result, care should be taken that plausibility is always preserved (i.e., that the interpretations do not run counter to the evidence, writ large). The limitations of textual evidence should also be borne in mind: Text can *only* tell us what people do (and not what they really think or feel). Any interpretations of the latter based on the former necessarily contain an element of speculation and risk being incorrect. At the same time, the researcher should try to construct the strongest possible evidential case for those interpretations she believes to be true.

What can be concluded about virtual community on the basis of the discourse evidence identified in the preceding sections? Specifically, what does CMDA reveal about the status of the two professional development sites as virtual communities? Our necessarily superficial analysis suggests some tentative interpretations. A close-to-the-data

interpretation would summarize the results given in the last paragraph of the preceding section: statements about the relative presence or absence of each of the community features analyzed in each of the two professional development environments. At this level, we might conclude that both environments manifest at least some of the hypothesized community behaviors. At the same time, differences exist in the degree to which each environment manifests the behaviors, and in which behaviors are manifested.

Our overall research question was: To what extent are these environments "virtual communities"? If "community" is operationalized according to the discourse behaviors in Table 5, and assuming for the sake of simplicity that all of the behaviors are equally indicative of "community" (a proposition open to debate), the Linguist List appears to be more community-like than the ILF, in that it manifests more community behaviors: presence of conflict and norms, and active, self-sustaining participation (in addition to the behaviors that the two environments share). Depending on our initial hypotheses, this result might be considered surprising: some theorists would predict that the ILF, as a multimodal environment, would create a richer social experience for users than a text-only environment (e.g., Media Richness Theory, Daft & Lengel, 1986). Moreover, the ILF was designed around a system of values (inquiry teaching) that its participants presumably share. How can we explain the greater evidence of community in the Linguist List?

The dimensions of variation summarized in Table 1 provide clues to interpretation. Listservs may be more effective at promoting professional development communities than Web sites, in that the former are "push" technologies and the latter "pull" technologies. Time being a resource in short supply for most teaching professionals, the convenience of receiving messages automatically (a medium variable) might make group members more likely to read and respond to them. The Linguist List also has a pre-existing offline community—professional linguists who meet face-to-face at conferences and read one another's work in professional journals, etc.—which provides (and sustains) a basis for online interaction. Regular off-line contact (a situational variable) may facilitate virtual community, raising levels of participant trust and emotional investment in the group.

Two other situational factors that conceivably facilitate the formation of virtual community are the fact that the Linguist List "owners" are peers in relation to the other participants (all are academic linguists), and that participants are free to select topics of discussion within the broad theme of academic linguistics, whereas on the ILF the "owners" and participants are in a hierarchical relationship (university professors and doctoral students vs. secondary school teachers and undergraduate teachers-in-training), and topics of discussion in the different areas of the site are more narrowly prescribed. A sense of shared ownership and empowerment to raise topics of discussion in an online environment may facilitate virtual community.[24] Additional analysis would be required to determine which of these factors is most explanatory.[25]

The question of whether the extent of community-like behavior is sufficient to justify labeling either environment a "virtual community" poses further interpretive challenges. How "community-like" must a group be in order to be a community? A researcher could establish objective criteria (e.g., certain key behaviors must be evident, or a certain combined frequency of a set of behaviors must be found), but this would necessarily be somewhat arbitrary. Ideally, such an assessment would take into account the perceptions of the participants themselves: it would hardly be satisfying to pronounce a

group a community on the basis of empirical discourse evidence, only to find that the participants themselves did not feel any sense of community-hood.[26]

At the broadest level, we might make theoretical interpretations about how the technological and social properties of CMC systems relate to the phenomenon of virtual community, extrapolating from the observations above. For example, we might use the comparison of the Linguist List and the ILF to argue against the Media Richness Theory (Daft & Lengel, 1986), since a lean, text-only environment was found to be more "community-like" than a rich, multimodal environment (cf. Walther, 1999). The properties of CMC systems also have practical implications for designers interested in creating environments to optimize community-like behavior. Designers need to be especially aware of the ways in which the features of such sites—e.g., push vs. pull message access, co-present vs. archived past messages, use of visual modalities such as video—encourage or discourage participation, arguably the *sine qua non* of community (Herring et al., 2002; but cf. Nonnecke & Preece, 2000). Finally, our CMDA analysis of virtual community necessitated the invention of new methods (e.g., coding categories) for identifying and quantifying communicative behaviors associated with virtual community. This is itself an original research contribution that could be refined and extended to other computer-mediated contexts in future studies.

The steps in the CMDA research process and their application to the problem of assessing the "virtual community" status of the two professional development groups are summarized in Table 6.

Conclusions

This chapter has presented a methodological overview of computer-mediated discourse analysis (CMDA), highlighting one empirical, linguistic approach.[27] This approach enables a level of empirical rigor, and reflects a heightened linguistic awareness, that sets it apart from other approaches to the study of Internet behavior. Five conceptual skills necessary for carrying out a CMDA analysis using the "code and count" method were discussed and applied to the concept of "virtual community," specifically the question of whether it exists in two asynchronous professional development environments. The existence of virtual community is a fundamental question that needs to be addressed if the term is to be used meaningfully, rather than purely metaphorically or (in Kling & Courtright's term) aspirationally, reflecting the user's desire that the positive aspects of community-hood be attributed to an online group.

Our hypothetical analysis suggested ways in which CMDA can shed empirical light on the notoriously slippery concept of virtual community. Crucially, CMDA requires that virtual community be operationalized according to behavioral criteria; on the Internet, such behavior takes place primarily through discourse. Although there is room for disagreement as to the best definition of virtual community, an operationalization need only be plausible and concrete in order to be applied and interpreted. Discourse measures are especially useful for comparing hypothesized community characteristics in different online environments or samples of data from the same environment. Further, once virtual community has been identified by discourse-independent means in some contexts, the discourse behaviors associated with it can be analyzed and extended as heuristics to identify virtual community in other contexts.

Table 6. Summary of the CMDA research process applied to a hypothetical question about virtual community

CMDA research process	Application to virtual community
Articulate research question(s)	E.g., "To what extent do two online professional development environments, listserv X and website Y, constitute "community"?"
Select computer-mediated data sample	E.g., intermittent time-based sampling (several weeks at a time at intervals throughout a year) of public messages from each group
Operationalize key concept(s) in terms of discourse features	Community → core participants + in-group language + support + conflict + group self-awareness + roles, etc.
Select and apply method(s) of analysis	Frequency counts of, e.g., messages and message length, rate of response ('core participants') Structural analysis of, e.g., abbreviations, word choice, language routines ('in-group language') Pragmatic analysis of, e.g., speech acts of positive politeness ('support'), etc.

Interpret results
  1. summarize/synthesize results of data analysis
  2. answer research question(s); explain unexpected results
  3. consider broader implications
	1. Listserv X has community features a, b, c, …; website Y has community features c, f, …
	2. Both have some community features; X is more community-like than Y. This is due to …
	3. Results have implications for: CMC theory (e.g., Media Richness); system design (e.g., push vs. pull access); research methodology (e.g., coding categories for community features)

In other respects, virtual community remains a challenging concept to demonstrate. Operationalizations are inevitably somewhat arbitrary; their value resides in being empirically testable, not in being true in an ultimate, philosophical sense. But what is virtual community, really? The concept is derivative of face-to-face community; thus a comparison between the two would seem to be logically required. However, CMC, by its very nature, arguably favors different kinds of group interactions than are possible face to face, causing other circumstances to vary in addition to the modality of the communication. Face-to-face community and online "community" may not be strictly comparable (Jones, 1995a); to what, then, can the latter be referenced to establish its existence? Moreover, the concept of "community" itself is inherently abstract, especially when stripped of its geographical basis, as is the case in "virtual" community. Whereas

certain behaviors, such as articulating norms and supporting others, might plausibly be associated with virtual community status, the same behaviors could also be interpreted in other ways, e.g., as power negotiation, or strategies to promote personal gain. That is, concluding on the basis of specific discourse features that a group is a virtual community might ultimately require too great an interpretive leap, given the abstractness of the target concept.

To a certain extent, these problems reflect the limitations of CMDA as an empirical, text-based approach. We can only directly analyze discourse behavior, and must infer larger social and cognitive formations (such as perceived group identity) indirectly. In fact, CMDA is most useful for comparing discourse features with independently established technical, social or psychological phenomena. Thus there are limits on what kinds of phenomena can be investigated via online discourse behaviors. However, this is also the case for self-report studies, ethnographic observation, social network analysis, and indeed for any other methodological approach to analyzing human behavior.

The coding and counting approach to CMDA illustrated in this chapter also has its strengths and limitations. The approach has the advantage of being based on a familiar social science paradigm, classical content analysis (Bauer, 2000), the usefulness of which has been repeatedly demonstrated for the analysis of communication media (Riffe et al., 1998; see also Bell & Garrett, 1998). It is particularly well-suited to analyzing and comparing discrete online phenomena, and for revealing systematic regularities in discourse use. However, quantitative content analysis may not be the best approach for analyzing complex, interacting, ambiguous or scalar phenomena, which risk distortion by being forced into artificially discrete categories for purposes of counting. Such phenomena may be more richly revealed by qualitative, interpretive approaches that illuminate through exemplification, argumentation and narration.[28]

The question then arises whether virtual community might more appropriately be analyzed by qualitative than by quantitative means. Its complexities and ambiguities have been illuminatingly discussed in ethnographic studies of recreational CMC environments by Baym (1999), Cherny (1999), Kendall (2002) and Reid (1991, 1994), among others. The ethnographic approach has been especially revealing in describing insider language use, rituals, norms and sanctions, and in narrating the histories of these practices. However, as Liu (1999) notes, most such studies assume *a priori* that the environments in question are communities (or in the case of Cherny, 'speech communities'), rather than assessing empirically the extent to which they meet a consistent set of criteria for community-hood. As a result, although ethnographic research can provide valuable insights into online environments in which participants may experience a strong sense of subjective belonging, the studies do not prove or disprove the existence of virtual community, nor can they be compared in any systematic way. It seems likely that both qualitative and quantitative approaches are needed in order to arrive at a full understanding of the nature of the online social groupings that currently proliferate in cyberspace.

At the same time, computer-mediated groups, including those that meet the subjective criterion of "feeling" like community to their members, are increasingly interacting via multimodal interfaces, including Web logs, online videoconferencing, and navigable virtual reality environments (Bowers, 2000; Kibby & Costello, 2001; Naper, 2001). The CMDA toolkit as articulated here is lacking in methods for analyzing meanings communicated through semiotic systems other than text. An important future direction for

CMDA is to identify and adapt appropriate methods of graphical, video and audio analysis to computer-mediated communication, on the assumption that these modalities communicate discourse meanings (Naper, 2001; Soukup, 2000; cf. Kress & van Leeuwen, 1996). With regard to online learning environments, Herring et al. (2002) have begun to do this in analyzing video clips on the ILF site; much more work in this direction remains to be done.

Acknowledgments

This chapter could not have been written without the feedback from the students in two graduate Computer-Mediated Discourse Analysis courses I taught at Indiana University in 2001-2002. I am grateful to those students for teaching me through their learning. Thanks are also due to Sasha Barab, Zilia Estrada and John Paolillo for helpful comments on the writing and organization of the chapter, and to Anthony Aristar for up-to-date information about the Linguist List. Any remaining errors of fact or interpretation are my own.

References

Alford, R. A. (1998). *The Craft of Inquiry: Theories, Methods, Evidence*. NY: Oxford University Press.

Atkinson, J. M., & Heritage, J. (Eds.) (1984). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Austin, J. L. (1962). *How to Do Things With Words*. Cambridge, MA: Harvard University Press.

Baker, A. (1998). Cyberspace couples finding romance online then meeting for the first time in real life. *CMC Magazine*, July 1998. Retrieved July 21, 2002 from the World Wide Web: http://www.december.com/cmc/mag/1998/jul/baker.html

Barab, S., MaKinster, J., Moore, J., Cunningham, D., & The ILF Design Team. (in press). Designing and building an online community: The struggle to support sociability in the Inquiry Learning Forum. *Educational Technology Research & Development*.

Bauer, M. (2000). Classical content analysis: A review. In M. Bauer & G. Gaskell (Eds.), *Qualitative Researching with Text, Image and Sound* (pp. 131-151). Thousand Oaks, CA: Sage.

Baym, N. (1995a). The emergence of community in computer-mediated communication. In S. Jones (Ed.), 138-163.

Baym, N. (1995b). The performance of humor in computer-mediated communication. *Journal of Computer-Mediated Communication*, 1 (2). Retrieved July 21, 2002 from the World Wide Web: http://www.usc.edu/dept/annenberg/vol1/issue2/baym.html

Baym, N. (1999). *Tune In, Log on: Soaps, Fandom, and Online Community*. Thousand Oaks, CA: Sage.

Bell, A., & Garrett, P. (Eds.) (1998). *Approaches to Media Discourse*. Oxford: Blackwell.

Bowers, J. (2000). *Weblog communities*. Posted February 28, 2000. Retrieved December 31, 2002, from the World Wide Web: http://irights.editthispage.com/stories/storyReader$115

Bruckman, A. S., & Resnick, M. (1995). The MediaMOO project: Constructionism and professional community. *Convergence*, 1 (1), 94-109.

Burkhalter, B. (1999). Reading race online. In M. Smith & P. Kollock (Eds.), 60-75.

Burnett, G. (2000). Information exchange in virtual communities: A typology. *Information Research*, 5 (4). Retrieved June 15, 2001 from the World Wide Web: http://www.shef.ac.uk/~is/publications/infres/paper82a.html

Chafe, W. (1994). *Discourse, Consciousness and Time*. Chicago: University of Chicago Press.

Cherny, L. (1999). *Conversation and Community: Chat in a Virtual World*. Stanford, CA: Center for the Study of Language and Information.

Cherny, L., & Weise, E. R. (Eds.) (1996). *Wired_Women: Gender and New Realities in Cyberspace*. Seattle: Seal Press

Condon, S. C., & Cech, C. (1996a). Discourse management strategies in face-to-face and computer-mediated decision making interactions. *Electronic Journal of Communication*, 6 (3). Retrieved June 15, 2001 from the World Wide Web: http://www.cios.org/www/ejc/v6n396.htm

Condon, S. C., & Cech, C. (1996b). Functional comparisons of face-to-face and computer-mediated decision making interactions. In S. Herring (Ed.), 65-80.

Condon, S. C., & Cech, C. (2001). Profiling turns in interaction: Discourse structure and function. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Los Alamitos: IEEE Computer Society.

Condon, S. C., & Cech, C. (in press). Discourse management in three modalities. In S. Herring (Ed.).

Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.

Daft, R. L., & Lengel, R. H. (1986). Organizational informational requirements, media richness and structural design. *Management Science,* 32, 554-571.

Danet, B., Ruedenberg-Wright, L., & Rosenbaum-Tamari, Y. (1997). Hmmm … where's that smoke coming from? Writing, play and performance on Internet Relay Chat. In S. Rafaeli, F. Sudweeks, & M. McLaughlin (Eds.), *Network and netplay: Virtual groups on the Internet* (pp. 41-76). Cambridge, MA: AAAI/MIT Press.

Diani, M. (2000). Social movement networks virtual and real. *Information, Communication & Society*, 3, 386-401.

Dibbell, J. (1993). A rape in cyberspace, or how an evil clown, a Haitian trickster spirit, two wizards, and a cast of dozens turned a database into a society. *Village Voice*, 36-42.

Etzioni, A. (1999). Face-to-face and computer-mediated communities, a comparative analysis. *The Information Society*, 15, 241-248.

Fairclough, N. (1992). *Discourse and Social Change*. London: Polity Press.

Fernback, J., & Thompson, B. (1995). Virtual communities: Abort, retry, failure? Retrieved December 31, 2002 from the World Wide Web: http://www.well.com/user/hlr/texts/VCcivil.html

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8 (1), 8-34.

Georgakopoulou, A. (in press). 'On for drinkies?': E-mail cues of participant alignments. In S. Herring (Ed.).

Glaser, B., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Press.

Goffman, E. (1959). *Presentation of Self in Everyday Life*. Garden City, NY: Anchor.

Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Gurak, L. (1996). The rhetorical dynamics of a community protest in cyberspace: What happened with Lotus Marketplace. In S. Herring (Ed.), 265-277.

Hall, K. (1996). Cyberfeminism. In S. Herring (Ed.), 147-170.

Harwood, W. S., Reiff, R., & Phillipson, T. (2001). *Conceptions of scientific inquiry: Voices from the front*. Unpublished ms, Indiana University, Bloomington.

Haythornthwaite, C., Kazmer, M. M., Robins, J., & Shoemaker, S. (2000). Community development among distance learners: Temporal and technological dimensions. *Journal of Computer-Mediated Communication*, 6 (1). Retrieved October 2, 2000, from the World Wide Web: http://jcmc.indiana.edu/vol6/issue1/haythornthwaite.html

Herring, S. C. (1992). Gender and participation in computer-mediated linguistic discourse. Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics. Document no. ED345552.

Herring, S. C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication, 3* (2). Retrieved June 15, 2001, from the World Wide Web: http://www.cios.org/www/ejc/v3n293.htm

Herring, S. C. (1994). Politeness in computer culture: Why women thank and men flame. In M. Bucholtz, A. Liang, L. Sutton, & C. Hines (Eds.), *Cultural performances*: *Proceedings of the Third Berkeley Women and Language Conference* (pp. 278-94). Berkeley, CA: Berkeley Women and Language Group.

Herring, S. C. (1996a). Posting in a different voice: Gender and ethics in computer-mediated communication. In C. Ess (Ed.), *Philosophical perspectives on computer-mediated communication* (pp. 115-145). Albany, NY: SUNY Press.

Herring, S. C. (1996b). Two variants of an electronic message schema. In S. Herring (Ed.), 81-106.

Herring, S. C. (Ed.) (1996c). *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (pp. 81-106). Amsterdam: John Benjamins.

Herring, S. C. (1998). Le style du courrier électronique: variabilité et changement. *Terminogramme, 84-85,* 9-16.

Herring, S. C. (1999a). Interactional coherence in CMC. *Journal of Computer-Mediated Communication*, 4 (4). Retrieved June 15, 2001, from the World Wide Web: http://jcmc.indiana.edu/vol4/issue4/

Herring, S. C. (1999b). The rhetorical dynamics of gender harassment on-line. *The Information Society*, 15 (3), 151-167.

Herring, S. C. (2001). Computer-mediated discourse. In D. Tannen, D. Schiffrin, & H. Hamilton (Eds.), *Handbook of Discourse Analysis* (pp. 612-634). Oxford: Blackwell.

Herring, S. C. (2002). Computer-mediated communication on the Internet. In B. Cronin (Ed.), *The Annual Review of Information Science and Technology* (pp. 109-168). Medford, NJ: Information Today Inc./American Society for Information Science and Technology

Herring, S. C. (in press a). Gender and power in online communication. In J. Holmes & M. Meyerhoff (Eds.), *Handbook of Language and Gender*. Oxford: Blackwell.

Herring, S. C. (in press b). Who's got the floor in computer-mediated conversation? Edelsky's gender patterns revisited. In S. Herring (Ed.).

Herring, S. C. (Ed.) (in press c). *Computer-mediated conversation*. Cresskill, NJ: Hampton Press.

Herring, S. C. (Under consideration). A classification scheme for computer-mediated discourse.

Herring, S. C., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society,* 18 (5), 371-383.

Herring, S. C., Johnson, D. A., & DiBenedetto, T. (1992). Participation in electronic discourse in a "feminist" field. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating power: The Proceedings of the Second Berkeley Women and Language Conference* (pp. 250-262). Berkeley, CA: Berkeley Women and Language Group.

Herring, S. C., Johnson, D. A., & DiBenedetto, T. (1995). "This discussion is going too far!" Male resistance to female participation on the Internet. In M. Bucholtz & K. Hall (Eds.), *Gender Articulated: Language and the Socially Constructed Self* (pp. 67-96). London: Routledge.

Herring, S. C., Martinson, A., and Scheckler, R. (2002). Designing for community: The effects of gender representation in videos on a Web site. *Proceedings of the 35th Hawaii International Conference on System Sciences*. Los Alamitos: IEEE Press

Herring, S. C., & Nix, C. G. (1997, March). *Is "serious chat" an oxymoron? Academic vs. social uses of Internet Relay Chat*. Paper presented at the annual meeting of the American Association of Applied Linguistics, Orlando, FL.

Hert, P. (1997). Social dynamics of an on-line scholarly debate. *The Information Society*, 13, 329-60.

Hiltz, R. S., & Turoff, M. (1978). *The Network Nation: Human Communication Via Computer*. New York: Addison-Wesley.

Jacobs-Huey, L. (in press). ...BTW, how do YOU wear your hair? Identity, knowledge and authority in an electronic speech community. In S. Herring (Ed.).

Jacobson, D. (1996). Contexts and cues in cyberspace: The pragmatics of naming in text-based virtual realities. *Journal of Anthropological Research*, 52, 461-481.

Jones, Q. (1997). Virtual communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer-Mediated Communication*, 3(3). Retrieved July 25, 2002 from the World Wide Web: http://jcmc.indiana.edu/vol3/issue3/jones.html

Jones, S. (1995a). Understanding community in the information age. In S. Jones (Ed.), 10-35.

Jones, S. (Ed.) (1995b). *Cybersociety: Computer-Mediated Communication and Community*. Thousand Oaks, CA: Sage.

Jones, S. (Ed.) (1998). *Cybersociety 2.0: Revisiting Computer Mediated Communication and Community*. Thousand Oaks, CA: Sage.

Kendall, L. (1996). MUDder? I hardly know 'er! Adventures of a feminist MUDder. In L. Cherny & E. R. Weise (Eds.), 207-223.

Kendall, L. (2002). *Hanging Out in the Virtual Pub*. Berkeley: University of California Press.

Kibby, M., & Costello, B. (2001). Between the image and the act: Interactive sex entertainment on the Internet. *Sexualities: Studies in Culture and Society*, 4 (3), 353-269.

Ko, K-K. (1996). Structural characteristics of computer-mediated language: A comparative analysis of InterChange discourse. *Electronic Journal of Communication*, 6 (3). Retrieved June 15, 2001, from the World Wide Web: http://www.cios.org/www/ejc/v6n396.htm

Kolko, B. (1995). Building a world with words: The narrative reality of virtual communities. *Works and Days*, 13 (1/2), 105-126. Retrieved June 15, 2001, from the World Wide Web: http://acorn.grove.iup.edu/en/workdays/toc.html

Kolko, B., & Reid, E. (1998). Dissolution and fragmentation: Problems in online communities. In S. Jones (Ed.), 212-229.

Korenman, J., & Wyatt, N. (1996). Group dynamics in an e-mail forum. In S. Herring (Ed.), 225-242.

Kress, G., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.

Lambiase, J. (in press). Hanging by a thread: Topic development and death in an electronic discussion of the Oklahoma City bombing. In S. Herring (Ed.).

Liu, G. Z. (1999). Virtual community presence in Internet Relay Chatting. *Journal of Computer-Mediated Communication*, 5 (1). Retrieved June 15, 2001, from the World Wide Web: http://jcmc.indiana.edu/vol5/issue1/liu.html

Livia, A. (in press). BSR ES TU F? Brevity and expressivity on the French Minitel. In S. Herring (Ed.).

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Longacre, R. E. (1992). The discourse strategy of an appeals letter. In W. Mann & S. A. Thompson (Eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text* (pp. 109-130). Amsterdam: John Benjamins.

Longacre, R. E. (1996). *The Grammar of Discourse*, 2nd ed. New York: Plenum.

Mann, C., & Stewart, F. (2000). *Internet communication and qualitative research: A handbook for researching online*. Thousand Oaks, CA: Sage.

McLaughlin, M. L., Osborne, K. K., & Smith, C. B. (1995). Standards of conduct on Usenet. In S. Jones (Ed.), 90-111.

Millen, D. (2000). Community portals and collective goods: Conversation archives as an information resource. *Proceedings of the 33rd Hawaii International Conference on System Sciences*. Retrieved June 15, 2001 from the World Wide Web: http://dlib.computer.org/conferen/hicss/0493/pdf/04933030.pdf

Murray, D. E. (1985). Composition as conversation: The computer terminal as medium of communication. In L. Odell & D. Goswami (Eds.), *Writing in Nonacademic Settings* (pp. 203-27). New York: Guilford.

Murray, D. E. (1988). The context of oral and written language: A framework for mode and medium switching. *Language in Society*, 17, 351-373.

Naper, I. (2001). System features of an inhabited 3D virtual environment supporting multimodality in communication. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Retrieved June 15, 2001 from the World Wide Web: http://www.hic.ss.hawaii.edu/HICSS_34/PDFs/DDPTC10.pdf

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*. D.C.: National Academy Press.

Nonnecke, B., & Preece, J. (2000). Persistence and lurkers in discussion lists: A pilot study. *Proceedings of the 33rd Hawaii International Conference on System Sciences*. Retrieved June 15, 2001 from the World Wide Web: http://dlib.computer.org/conferen/hicss/0493/pdf/04933031.pdf

Panyametheekul, S. (2001). Disrupted adjacency and cohesion in Thai chat. Unpublished ms, Indiana University, Bloomington.

Paolillo, J. (1996). Language choice on soc.culture.punjab. *Electronic Journal of Communication*, 6 (3). Retrieved June 15, 2001, from the World Wide Web: http://www.cios.org/www/ejc/v6n396.htm

Paolillo, J. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics*, 5 (2), 180-213.

Paolillo, J. (in press). Conversational codeswitching on Usenet and Internet Relay Chat. In S. Herring (Ed.).

Psathas, G. (1995). *Conversation Analysis: The Study of Talk-in-Interaction*. Thousand Oaks, CA: Sage.

Rafaeli, S., & Sudweeks, F. (1997). Networked interactivity. *Journal of Computer-Mediated Communication*, 2 (4). Retrieved June 15, 2001, from the World Wide Web: http://jcmc.indiana.edu/vol2/issue4/

Ravert, R. (2001.) Adolescent chat style. Unpublished ms, Indiana University, Bloomington.

Reid, E. M. (1991). *Electropolis: Communication and Community on Internet Relay Chat*. Senior Honours thesis, University of Melbourne, Australia. Retrieved December 31, 2002, from the World Wide Web: http://www.aluluei.com/

Reid, E. M. (1994). *Cultural Formations in Text-Based Virtual Realities*. Unpublished master's thesis, University of Melbourne, Australia. Retrieved June 15, 2001, from the World Wide Web: http://home.earthlink.net/~aluluei/cult-form.htm

Reid, E. M. (1998). Hierarchy and power: Social control in cyberspace. In M. Smith & P. Kollock (Eds.), 107-133.

Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley. Retrieved June 15, 2001, from the World Wide Web: http://www.rheingold.com/vc/book/

Riffe, D., Lacy, S., & Fico, F. (1998). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Hillsdale, NJ: Erlbaum.

Sacks, H. (1984). On doing "being ordinary". In J. M. Atkinson & J. Heritage (Eds.), 413-429.

Severinson Eklundh, K. (1986). *Dialogue Processes in Computer-Mediated Communication: A Study of Letters in the COM system*. Linköping Studies in Arts and Sciences 6. University of Linköping.

Smith, M. & Kollock, P. (Eds.) (1999). *Communities in Cybespace*. London: Routledge.

Soukup, C. (2000). Building a theory of multi-media CMC. *New Media & Society*, 2, 407-425.

Swales, J. (1990). *Genre Analysis*. Cambridge: Cambridge University Press.

Tannen, D. (Ed.) (1993). *Framing in Discourse*. NY: Oxford University Press.

Voth, C. (1999). *The Facts on FAQs: Frequently Asked Questions Documents on the Internet and Usenet*. Unpublished master's thesis, University of Texas at Arlington.

Walther, J. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, 23 (1), 3-43.

Walther, J. (1999). Visual cues and computer-mediated communication: Don't look before you leap. Retrieved December 31, 2002, from the World Wide Web: http://www.rensselaer.edu/~walthj/ica99.html

Weber, H. L. (in press). Missed cues: How disputes can socialize virtual newcomers. In S. Herring (Ed.).

Wellman, B. (2001). Message posted to AIR-L@ , December 25, 2001.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Werry, C., & Mowbray, M. (Eds.) (2001). *Online Communities: Commerce, Community Action, and the Virtual University*. Upper Saddle River, NJ: Prentice Hall.

Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing. In S. Herring (Ed.), 29-46.

Notes

[1] See, e.g., Burnett (2000), who characterizes "virtual communities" broadly as "discussion forums focusing on a set of interests shared by a group of geographically dispersed participants." According to this characterization, almost any Internet discussion group is a virtual community.

[2] For examples of this usage, see Ferrara et al. (1991), who employ the term 'register' in this broad sense, and, more recently, Crystal (2001), who refers to the language of the Internet as 'netspeak.'

[3] 'Textual' is intended here broadly, to include any form of language, spoken or written, that can be captured and studied in textual form.

[4] For a relatively current discussion of ethical issues associated with collecting and analyzing data from the Internet (although as of this writing, understandings of what is acceptable practice are still evolving), see Mann and Stewart (2000).

[5] Gathering and comparing evidence from multiple analytical approaches is known as triangulation.

[6] The Linguist List has subsequently expanded its Web presence, coming to serve as an electronic clearing house for language- and linguistics-related resources.

[7] This strategy was adopted, for example, by Herring (1992, 1993).

[8] This question assumes a common set of criteria for both domains, and the availability of data for face-to-face communities.

[9] Cells above the double line in Table 1 indicate medium (technological) variables; cells below the double line indicate situational (social) variables (see Herring, u.c. for a full description of this system of classification).

[10] Causal indeterminacy in CMDA research can be minimized in two ways. First, data samples that are more similar than different can be selected, in an attempt to approximate the experimental approach of holding all but one feature constant. Second, dimensions of variation within the data sample(s) can be considered in interpreting the research findings (see Herring, u.c., and 'interpretation' below). In some cases, although differences could result in principle from multiple contrasting dimensions, in practice, the evidence points more strongly to one than to the others.

[11] In 1997, Linguist made available a new distribution option, Linguist Lite, which sends subscribers a single message containing the subject headers of the day's messages; subscribers must then go to the Linguist Web site to read the messages. This option, which exists alongside the traditional listserv distribution format, combines both "push" and "pull" elements.

[12] Even then, this method is likely to produce more data than can reasonably be analyzed using most linguistic methods, such that further winnowing of the sample may be required.

[13] Among the advantages of ongoing observation is that it allows the researcher to capture data opportunistically, should interesting interactions take place outside the formally established data collection periods.

[14] For example, chi-squared tests, which compare actual with expected distributions of results, typically require a minimum of five instances in each sub-category.

[15] For one thing, people can engage in large group conversations online, whereas a conversation involving one hundred or more people would be impossible face-to-face (Herring, 1999a).

[16] In their study of participation in the video-centered 'classroom' discussions on the ILF, Herring et al. (2002) found that male in-service teachers featured in the videos, and female ILF development team members, were the most active participants, suggesting that both status and gender are associated with level of engagement in the site.

[17] The criterion of research reproducibility has traditionally been a guiding force in scientific methodology (cf. Swales, 1990).

[18] For an alternative set of criteria, and an attempt to operationalize them empirically, see Liu (1999), who bases his analysis of community in Internet Relay Chat (IRC) on Jones' (1997) four criteria for a "virtual settlement:" (1) a virtual common-public-space; (2) a variety of communicators; (3) a minimum level of sustained stable membership; and (4) a minimum level of interactivity.

[19] The Linguist List has many international subscribers, but most messages are posted in English, the international language of scholarship.

[20] There are several possible reasons why the Linguist List is more conflict-prone than the ILF, despite the fact that the former is moderated and the latter is not. The Linguist List is larger and more impersonal than the ILF, which has restricted membership and makes available individual user profiles. Linguist messages are archived out of sight, while ILF messages remain on the site. The professional discourse of academic linguists is also probably more antagonistic than that of secondary school teachers in off-line contexts. Social accountability, message persistence, and generally supportive professional norms of communication could inhibit criticism and conflict in postings to the ILF. Alternatively, it could be that ILF participants are not as engaged in their interactions as are Linguist List participants.

[21] While some of these phenomena are conventionally associated with particular linguistic means of expressions (e.g., "Thanks" and "I'm sorry" as expressions of politeness), they can also be expressed indirectly or unconventionally (e.g., "That's sweet of you" and "What a klutz I am"). Given the creativity of language users, it is nearly impossible to predict in advance what all the variants might be.

[22] This need not be a problem, provided enough data are analyzed to meet the criterion of sufficient data to run tests of statistical significance, as noted in the section on "data". If structural and semantic analyses are conducted of the same data sample, it is possible to code all of the data for the relevant structural phenomena, and a selected sub-set of the data for the semantic phenomena

[23] Knowledge tends to be expressed as opinions on the Linguist List (Herring, 1996b), and as advice and personal experience on ILF (Herring et al., 2002).

[24] Cf. Bruckman & Resnick's (1995) suggestion that "letting the users [of a professional development MOO] build a virtual world rather than merely interact with a pre-designed world gives them an opportunity for self expression, encourages diversity, and leads to a meaningful engagement of participants and enhanced sense of community."

[25] One direction such analysis might take would be to hypothesize that a given difference is especially significant, and analyze new data samples that vary only (or predominantly) according to that dimension. For example, two web-based forums targeting similar

audiences for similar purposes, one created and maintained on a volunteer basis by peers, and the other created and controlled by "experts", could be compared for evidence of community behaviors to test the hypothesis that a sense of "shared ownership" facilitates virtual community. Another possibility would be to conduct multivariate analyses on a large number of samples that vary according to multiple dimensions.

[26] Conversely, participants might experience a sense of belonging and identity even in groups where discourse behaviors associated with community are lacking. For example, Nonnecke and Preece (2000) interviewed "lurkers" in online discussion groups and found that some expressed a sense of belonging, even though they never posted messages to the group.

[27] Interpretive approaches to CMDA, drawing on methods from, e.g., anthropology and rhetoric, also exist. See, for example, Cherny (1999) and Kendall (2002) for anthropological (ethnographic) approaches; Gurak (1996) and Herring (1999b) for rhetorical approaches.

[28] Qualitative approaches fall within the purvue of CMDA, provided they are based on analysis of actual records of online interaction. Examples of qualitative CMDA research, in addition to those mentioned in note 26, include Baym (1995b); Danet et al. (1997); Herring, Job-Sluder, Scheckler & Barab (2002); Livia (in press); and Weber (in press). Moreover, even rigorously quantitative CMDA analysis can benefit from a theoretically-informed interpretive framework, "thick" description of users, systems and contexts, and discourse examples to lend analytical nuance.