

A Faceted Classification Scheme for Computer-Mediated Discourse

Susan C. Herring

Indiana University, Bloomington

Abstract

This article describes a classification scheme for computer-mediated discourse that classifies samples in terms of clusters of features, or “facets”. The goal of the scheme is to synthesize and articulate aspects of technical and social context that influence discourse usage in CMC environments. The classification scheme is motivated, presented in detail with support from existing literature, and illustrated through a comparison of two types of weblog (blog) data. In concluding, the advantages and limitations of the scheme are weighed.

1. Introduction

It is by now a truism that computer-mediated communication (CMC) – defined here as predominantly text-based human-human interaction mediated by networked computers or mobile telephony – provides an abundance of data on human behavior and language use. Confronted with such abundance, researchers and practitioners have naturally sought to group, label, or otherwise organize CMC into categories that would facilitate its analysis and uses. However, there has been neither systematic discussion of how this should be done nor consensus regarding individual attempts to do so, many of which have been implicit and ad hoc. As a consequence, how to classify CMC remains a significant unaddressed problem of information organization.

This article is concerned with the classification of CMC for research purposes, with a focus on online language and language use, hereafter referred to as computer-mediated discourse (CMD; Herring 1996, 2001). Specifically, it proposes an approach to the classification of CMD based on multiple categories or “facets”, a concept borrowed from classification theory in the field of library and information science. In contrast to applications in that field, however, which are primarily concerned with information storage and retrieval, the goal of the CMD scheme is to articulate aspects of context – both technical and social – that potentially influence discourse

usage in CMC environments, and thereby to bring them to the conscious attention of the researcher. In this, it is akin in spirit to Hymes' (1974) etic grid, also known as the SPEAKING mnemonic, which is treated here as an early example of faceted classification in a research context.

The organization of this article reflects its goal to motivate, articulate, and illustrate a model. The next section identifies the basic problem that gave rise to the need for a CMD classification scheme. Following a review of research on discourse classification, I then present an overview of the proposed faceted classification scheme for CMD and describe its dimensions and categories. This is followed by an illustration in which the scheme is applied to characterize contrasting computer-mediated (weblog) data samples. In concluding, the advantages and limitations of the faceted classification approach to online communication are weighed.

1.1 The problem

Various attempts have been made by linguists to classify CMD, starting in the 1980s and early 1990s. Accustomed to dealing with two basic *modalities* of language – speech and writing – these linguists first asked: Is it a type of writing, because it is produced by typing on a keyboard and read as text on a computer screen? Is it “written speech” (Maynor 1994), because it exhibits features of orality, including rapid message exchange, informality, and representations of prosody? Or is it a third type, intermediate between speech and writing, or in any event characterized by unique production and reception constraints (Ferrara, Brunner & Whittmore 1991; Murray 1990)?

These early efforts at classification tended to overgeneralize about computer-mediated language, as if CMD were a single, homogeneous *genre* or communication type. Even in recent years, “Netspeak” has been posited as an emergent, global variety of online language characterized by abbreviations, emoticons, and nonstandard spellings (Crystal 2001).

However, as awareness of CMC spread with the popularization of the Internet, it soon became apparent that computer-mediated discourse was sensitive to a variety of technical and situational factors, making it complex and variable (Baym 1995; Cherny 1999; Herring 1996). Simultaneously, the focus of much CMD research shifted to describing the linguistic features of individual *genres of CMD*, e.g., email discussion lists, Usenet newsgroups, Internet Relay Chat (IRC), and MUDs.¹ Elsewhere, I have termed these “socio-technical modes” (Herring 2002) – following Murray’s (1988) use of the term “mode” to refer to technologically-defined CMC subtypes – to reflect the fact that labels such as “IRC”, “Usenet”, “email”, and so forth are commonly understood to refer not just to CMC systems, but also to the social and cultural practices that have arisen around their use.

The genre and mode approaches, however, while preferable to lumping all CMC into a single type, are also limited as a basis for classification of CMD. First, the concept of genre can potentially be applied to communication at different levels of specificity (Maingueneau 1998), and is thus imprecise. For example, is the appropriate level of genre classification “email discussion lists”, “academic discussion lists” (cf. Grüber 2000) or “academic discussion lists on masculine/feminine topics” (cf. Herring 1996) – each of which is associated with characteristic linguistic practices? The mode approach partially addresses this problem, in that it refers primarily to technologically-defined CMD types,² but it neglects social distinctions of the sort identified by Grüber (2000) and Herring (1996).

Another limitation of both the genre and mode approaches is that they are most easily applied to classify discourse that takes place using established, named technologies (cf. Swales 1990), such as those that are popular on the Internet. It is less clear how either approach could be

¹ See, for example, Werry (1996) for IRC, Baron (1998) for email, Cherny (1999) for social MUDs, and Grüber (2000) for academic discussion lists.

² In the case of the example of email-based discussion, “listservs” are a mode, as distinct from “newsgroups” and “Bulletin Board Systems (BBS)”, based first and foremost on their different technical configurations (e.g., push vs. pull delivery; subscription/registration requirements).

used to classify new and emergent forms of CMD, or discourse that takes place via customized systems that operate within restricted (e.g., educational, governmental, organizational) domains. For these, a more flexible classification system is needed.

The approach to the classification of computer-mediated discourse proposed in this article is based on multiple categories or “facets”. These categories cut across the boundaries of socio-technical modes, and combine to allow for the identification of a more nuanced set of computer-mediated discourse types, while avoiding the imprecision associated with the concept of genre. Since the classification scheme does not rely on pre-existing modes, it can also be applied to discourse mediated by emergent and experimental CMC systems. The scheme is intended primarily as a faceted lens through which to view CMD data in order to facilitate linguistic analysis, especially research conducted in the discourse analysis, conversation analysis, pragmatics, and sociolinguistics traditions.³ It is intended to complement genre or mode-based analyses, which can provide a convenient shorthand for categorizing CMD types, but are less precise and flexible.

2. Background

2.1 Conceptual foundations

The CMD classification scheme is a core component of the computer-mediated discourse analysis (CMDA) approach developed by Herring (2001, 2004a);⁴ the scheme is presented here in detail for the first time. CMDA adapts methods from the study of spoken and written discourse to computer-mediated communication data. Similarly, the central role of classification in CMDA can be traced back to traditional discourse analytic concerns.

Discourse analysts have traditionally classified discourse into types according to various criteria. These include *modality*, *number of discourse participants*, *text type* or *discourse type*, and

³ For a recent overview of research in the sociolinguistics tradition, see Androutsopoulos (2006).

⁴ The other core components of CMDA are *levels of analysis and operationalization of concepts*; see Herring (2004a).

genre or *register* (table 1). While the definitions and boundaries of these distinctions have been much debated, they can be understood as being in a generally non-exclusive and hierarchical relationship to one another (e.g., casual chat is a type of conversation, typically a dialogue and typically produced via speech). As noted above, however, genre can be analyzed on multiple levels of generality, and thus all of the types in table 1 have also been characterized as “genres”.⁵ Further, Biber (1988) has challenged the validity of the spoken/written language distinction, proposing that discourse types be situated instead along multiple continua.

Classification criteria	Types	Invoked by
Modality (means of production/reception)	speech, writing	Chafe and Danielewicz (1987)
Number of discourse participants	monologue, dialogue, polylogue	Dooley and Levinsohn (2001)
Text/Discourse type	conversation, narrative, exhortation, exposition, etc.	Longacre (1996); Virtanen (1992)
Genre/Register ⁶	casual chat, interview, public lecture, personal letter, short story, scientific research article, etc.	Biber (1988); Swales (1990)

Table 1. Traditional approaches to discourse classification

Despite their disagreements, discourse analysts implicitly agree that classification facilitates analysis. This is because exemplars of the same type of discourse tend to share features

⁵ It is also possible to identify sub-genres of the genres in table 1, for example, a job interview as compared to an interview on a radio or television talk show, a personal Christmas letter as compared to a personal letter breaking off relations with one’s paramour (i.e., a *Dear John* letter).

⁶ In the sense of Biber (1988). “Register” has another usage in linguistics (as a shorthand for formal/informal style) that is not intended here.

that distinguish them collectively from other discourse types; classification makes this explicit, thereby facilitating comparison across types.

Classification may also serve to remind the analyst to attend to important properties of the data under consideration, even when no overt comparison is involved. For example, spoken discourse typically has shorter sentences and words, more sentence fragments, and more markers of interpersonal relations than discourse produced in writing (Chafe & Danielewicz 1987). A researcher interested in studying sentence complexity might analyze both spoken and written texts, but to do so without taking modality into account could result in overlooking systematic, conditioned patterns in the data. Moreover, certain linguistic and rhetorical phenomena occur regularly only in certain discourse or text types. Examples include turn taking in spoken dialogue, plot development in narrative, and argumentation in expository discourse (Longacre 1996; Virtanen 1992). A researcher interested in turn taking, for example, must identify text type as a precursor to further linguistic analysis.

A different approach sometimes adopted in spoken discourse classification is the ethnography of communication model of Dell Hymes (1974), reproduced in figure 1.

Setting/Scene	The setting refers to the time and place, while scene describes the “psychological setting” or “cultural definition” of a scene.
Participants	Speaker and audience.
Ends	Purposes, goals, and outcomes.
Act sequence	Form and order of events.
Key	The “tone, manner, or spirit” of the speech.
Instrumentalities	Channels, forms, and styles of speech.
Norms	Social rules governing the event and the participants’ actions and reaction.
Genres	The type of speech or event.

Figure 1. The SPEAKING model (Hymes 1974)

Hymes' taxonomy comprises the categories Setting/Scene, Participants, Ends, Act sequence, Key, Instrumentalities, Norms, and Genres, which together form the acronym SPEAKING. This model has been widely applied to characterize novel or exotic speech communities (e.g., Nevins 2004), serving as what Hymes calls an "etic grid", or preliminary descriptive framework, that draws the researcher's attention to aspects of the speech situation that may assist in interpreting linguistic phenomena of interest.

Analysts of computer-mediated discourse have many of the same needs for classification as traditional spoken and written discourse analysts: Properties of the medium that predict language variation must be identified; CMD modes must be characterized, and novel CMD situations call for etic description. These needs are compounded by the rapid pace with which new computer-mediated communication technologies, such as SMS (text messaging through mobile phones), instant messaging, and blogs, have emerged into popular use over the past decade (Herring 2004b). Other technologies will inevitably follow, placing a continuing demand on linguists to provide systematic, meaningful characterizations of discourse in emergent mediated environments.

2.2 Previous classification of CMD

Three approaches can be distinguished in efforts to classify computer-mediated discourse to date. As noted at the outset, a number of early researchers sought to characterize computer-mediated discourse as a whole, often based on limited data.⁷ Ferrara et al. (1991), for example, described CMD as an "emergent register" based on their study of one type of experimental, synchronous CMD. Crystal's (2001) characterization of the language of the Internet as "Netspeak" is a more recent example of this globalizing approach. Relatedly, early attempts to classify CMD in relation to speaking and writing tended to consider only one form of CMD (Werry 1996; Yates 1996), although some researchers have suggested a continuum along which asynchronous CMD occupies

⁷ Notable exceptions are Murray (1988) and Severinson-Eklundh (1986).

Language@Internet 1/2007 (<http://www.languageatinternet.de>, urn:nbn:de:0009-7-7611, ISSN 1860-2029)

a position closer to writing, and synchronous CMD occupies a position closer to speaking (e.g., Herring 2001).

Later researchers narrowed their focus of attention to individual modes of CMD, describing the characteristics of communication in each.⁸ An example of this approach from a linguistic perspective is Cherny's (1999) extended ethnographic study of a social MUD. Cherny (1999) emphasized that the norms for discourse in a social MUD are not the same as those for Internet Relay Chat, despite the fact that both are synchronous chat environments. Linguistic variation can be observed between one social MUD and another, based on the histories, norms, and user demographics of each group, leading Cherny to characterize individual MUDs as "speech communities".

The third approach, which most closely resembles that taken in the present article, involves classifying CMD data according to a pre-defined set of categories. As early as 1988, Murray applied a Hymesian grid to characterize different forms of CMD in use among workers in a large U.S. technology organization. Collot and Belmore (1996) also adopted Hymes' taxonomy to describe asynchronous BBS data, as a preliminary to quantitative analysis. Although their focus was not on language, Rice and Gattiker (2000) developed an extensive classification grid in which they situated CMC in relation to other forms of mediated communication. However, they did not justify the construction of the grid, nor apply it to data analysis.

In her analysis of television soap opera fan newsgroups, Baym (1995: 141) drew on previous research to identify five factors that condition variation in CMD: the external contexts – physical, cultural, and subcultural – in which CMC use is situated; the temporal structure of the group; the computer system infrastructure; the purpose of communication; and the characteristics of the group and its members. Baym's approach has a number of advantages: It is grounded in empirical observations; it is tailored to CMD data and takes the contributions of the computer

⁸ For an overview of this research, see Herring (2002).

system into account; and its utility has been demonstrated through application to data. A disadvantage is that it is limited to only five factors; it does not include, for instance, the languages of the participants or the fonts available to express them (cf. Danet & Herring 2007).

In none of the studies mentioned above was classification the primary objective. Rather, CMD researchers have characterized their data in pursuit of other goals, to distinguish them from other kinds of data, and to invoke factors that explain their characteristics. The goal of the present article is to systematize and extend these efforts in a classification scheme intended to highlight those features of CMC that most directly affect users' linguistic choices.

3. Faceted classification

Faceted classification is an approach to the organization of information with origins in the field of library and information science. First systematized as a science by Ranganathan (1933) to classify books in libraries, it was later developed by the *U.K. Research on Classification Group* (Vickery 1960) for the organization of document collections in scientific fields, where it proved effective in the storage and retrieval of compound and complex subjects. More recently, faceted classification has been implemented to assist automated search and retrieval of information (Prieto-Diaz 1991), including on the Web (Broughton & Lane 2000), and has been extended to other fields and knowledge domains (e.g., art and architecture; Tudhope et al. 2002).

Facets are categories or concepts of the same inherent type. A faceted scheme has several facets and each facet may have several terms, or possible values, e.g., a faceted classification scheme for wine might include the facets (and terms) "grape varietal" (riesling, cabernet sauvignon, etc.), "region" (Napa Valley, Rhine, Bordeaux, etc.), and "year" (2001, 2002, etc.). Ranganathan (1933) described the faceted classification method as analytico-synthetic: A subject domain is first analyzed into component facets, and relevant facets are then synthesized into combinations to characterize items of interest. Thus many facets may be applied to the description of wine, but only a subset of them – such as varietal and region – may be relevant to classifying

wines for the purpose of marketing them to casual consumers. The flexibility of faceted classification lies in its ability to describe a large number of items within the subject domain, including novel items, on the basis of a relatively economical, pre-defined set of facets and terms. The facets need not be ordered, nor be of the same type, although they should be clearly defined and mutually exclusive.

The present model involves faceted classification in the general sense described above, although it does not adhere to the specific criteria laid out by Ranganathan (1933) and others regarding selection of facets for a given subject area. This is in part because the CMD scheme was not designed from the top down as a faceted classification scheme, but rather evolved from the bottom up, as in the case of Baym's (1995) five factors that condition variation in CMC. Moreover, as noted at the outset, its purpose is not to facilitate information storage and retrieval, but rather to facilitate data selection and analysis in CMD research. These differences aside, the CMD scheme functions in many ways like a traditional faceted classification scheme, and has similar advantages and limitations.

4. The faceted classification scheme for CMD

4.1 Overview of the faceted classification scheme

The classification approach to CMD presented here is organized at the highest level by the assumption that computer-mediated discourse is subject to two basic types of influence: medium (technological) and situation (social). These are presented in an unordered, non-hierarchical relationship, on the further assumption that one cannot be assigned theoretical precedence over the other for CMD as a whole; rather, the relative strength of social and technical influences must be discovered for different contexts of CMD through empirical analysis.

Under each influence type, a number of categories (facets) are posited, along with several possible realizations (terms) for each. The categories were arrived at in an inductive manner on the basis of empirical evidence from the CMD research literature in answer to the question: What

factors condition variation in computer-mediated language use? The proposed scheme is a preliminary attempt to aggregate and classify this body of knowledge.

The first set of categories describes technological features of computer-mediated communication systems. These are determined by messaging protocols, servers and clients, as well as the associated hardware, software, and interfaces of users' computers, in as much as it is possible for the researcher to obtain such information. The inclusion of a set of technological factors in the approach does not assume that the computer medium exercises a determining influence on communication in all cases (a position known as technological determinism, cf. Markus 1994), although each factor has been observed to affect communication in at least some instances. One reason for including *medium factors* as a separate set is, precisely, to attempt to discover under what circumstances specific system features affect communication, and in what ways.

The second set consists of social factors associated with the situation or context of communication. These include information about the participants, their relationships to one another, their purposes for communicating, what they are communicating about, and the kind of language they use to communicate (cf. Baym 1995; Hymes 1974). The inclusion of a set of *situation factors* assumes that context can shape communication in significant ways, although it does not assume that any given factor is always influential. The particular factors included in the model described below have all been observed to condition variation in at least some CMD contexts.

As in traditional faceted classification, these two sets of categories are open ended; additional factors can be added as justified by evidence that they affect online discourse. Also, within each set, the categories are unordered and not assumed *a priori* to be in any particular relationship to one another. Categories may (or may not) interact, just as there may (or may not) be patterned correspondences between medium and situation factors, in principle. In fact, modes

of CMD such as “listserv lists” and “Internet Relay Chat” exhibit characteristic combinations of facets, as discussed further below.

The categories themselves are each realized by more than one possible value. As in traditional faceted classification, the categories may be heterogeneous, with values that are binary (e.g., message transmission = 1-way or 2-way), scalar (e.g., degree of persistence of text = low → high), or a list of discrete items (e.g., topic = Chinese restaurants in Paris; last presidential elections; marsupials; etc.); the latter type may be open ended.

The most straightforward procedure for applying the scheme is as follows. Once a sample or corpus of CMD has been identified, the researcher goes through the categories for each set, assigning the appropriate value for each category based on the information available to him or her from the data, additional contextual knowledge he or she may possess, or general knowledge of CMC. One or more categories may not be applicable to a particular CMD sample, in which case no value is assigned for them.

This process should produce a list of all applicable values for the categories in each set. The researcher may then select from the list of values those that are relevant to his or her analytical purposes. In this sense, the scheme is analytico-synthetic (cf. Ranganathan 1933). As in traditional faceted classification, it is also possible to apply the scheme selectively, by assigning values only to those categories or facets that are relevant to the analysis.

The scheme may be applied to data samples of almost any size, although not all categories are relevant for very small samples. For example, a sample of a single message does not readily allow for generalizations about the “group” of which it is a part. Conversely, very large samples may contain so much internal variation that it is meaningless to assign a single value for each feature. In such cases, multiple values may be assigned to a feature for purposes of overall characterization. The researcher may also decide to apply the scheme at the level of contrasting sub-samples in order to better characterize their distinguishing properties.

4.2 Medium factors

This section and the following section enumerate and define the categories of the CMD classification scheme and cite empirical studies to justify their inclusion. The citations are meant to be indicative only; many other studies could be cited that contribute relevant evidence.

Table 2 lists some of the most important *medium* factors that have been observed to condition computer-mediated discourse, and that are therefore posited as categories in the classification scheme. Although they are not in any necessary order, they are numbered in table 2 for ease of reference.

M1	Synchronicity
M2	Message transmission (1-way vs. 2-way)
M3	Persistence of transcript
M4	Size of message buffer
M5	Channels of communication
M6	Anonymous messaging
M7	Private messaging
M8	Filtering
M9	Quoting
M10	Message format

Table 2. Medium factors

The first medium factor relates to *synchronicity* of participation (Kiesler, Siegel & McGuire 1984). Asynchronous systems do not require that users be logged on at the same time in order to send and receive messages; rather, messages are stored at the addressee's site until they can be read. Email is an example of this type. In synchronous systems, in contrast, sender and addressee(s) must be logged on simultaneously; various modes of "real-time" chat are the most

common forms of synchronous CMC.⁹ Most traditional forms of writing are asynchronous, and spoken conversation is typically synchronous, making synchronicity a useful dimension for comparing different types of CMC with spoken and written discourse (Condon & Cech 1996; Ko 1996; Yates 1996). Synchronicity is also a robust predictor of structural complexity, as well as many pragmatic and interactional behaviors, in computer-mediated discourse (Herring 2004a; Ko 1996).

A cross-cutting technological dimension has to do with the granularity of the units that are transmitted by the CMC system, that is, whether the transmission is message-by-message, or character-by-character (a third possibility is line-by-line transmission). This has implications for whether or not simultaneous feedback is available during message exchange. With message-by-message transmission, the receiver does not typically have any indication that the sender is composing a message until it is sent and received;¹⁰ thus, it is impossible for the receiver to interrupt or otherwise engage simultaneously with the sender's message. Cherny (1999) terms this transmission "one-way"; most CMC systems in current use make use of one-way transmission.

In contrast, character-by-character transmission is "two-way", in that both the sender and the receiver are able to see the message as it is produced, making it possible for the receiver to give simultaneous feedback. In two-way CMC systems, participants' screens split into two (sometimes more) parts, and the words of each participant appear keystroke-by-keystroke in their respective parts as they are typed. Examples of two-way synchronous CMC include the VAX "phone" protocol studied by Anderson, Beard and Walther (forthcoming), UNIX "talk", and the split-screen mode of ICQ (Herring 2002). Anderson, Beard, and Walther (forthcoming) have observed that two-way transmission can profoundly alter the structure of turn taking.

⁹ CMC systems of intermediate synchronicity also exist; for example, Babble (Erickson et al. 1999), an experimental chat-like system with a scroll-back log that persists for days, allows users who missed real-time messages to read them later. Instant messaging clients similarly blur the boundary by allowing users to read messages sent while they were away from their computer upon their return, as long as their IM client remains open.

¹⁰ An exception is instant messaging systems that indicate that a participant is typing a message, without yet displaying what is being typed.

“Persistence of transcript” refers to how long, relatively speaking, messages remain on the system after they are received. Email is persistent by default, remaining in users’ mail queues or files until deleted by the users. Moreover, many listservs archive email messages sent to discussion lists, and messages posted to Usenet newsgroups have been archived since 1995 (first by dejanews.com, and since 2000, by Google). In contrast, most chat systems retain only a few screens of messages in their scrollbar buffer, with old messages eventually disappearing as they are replaced by new ones. Even the messages in the buffer disappear when the user ends a chat session, unless he or she has chosen to log the interaction. Thus, chat is relatively ephemeral compared to email, but it is more persistent than spoken conversation, in that one’s typed words linger before they scroll out of sight. The overall greater persistence of CMD heightens meta-linguistic awareness: It allows users to reflect on their communication – and play with language – in ways that would be difficult in speech. It also allows them to keep track of, and participate in, multiple conversational threads (Herring 1999).

“Size of message buffer” refers to the number of characters the system allows in a single message. In most email-based systems, the buffer is effectively limitless – or at least, it is larger than practical limits on how long most people are willing to type and others are willing to read. Many chat systems, however, impose limits on message size, and text messaging systems on mobile telephones limit users to 160 characters per message. Condon and Cech (2001) found that smaller buffers often mean shorter messages and different discourse organizational strategies (see also Baron forthcoming); small buffers also increase the likelihood that language will be structurally abbreviated (Anis 2007).

With multimedia increasingly augmenting textual online interaction, it is important to take into account how many and what kinds of “channels of communication” a CMC system makes available. Visual channels in addition to text include graphics (static or animated) and video; videoconferencing systems (such as CUseeMe and audiochat; Chou 1999) provide an audio

channel as well. Herring, Martinson, and Scheckler (2002) found that the presence and content of video images affected the amount and gender distribution of discourse on an educational website. Communication involving Voice-over-Internet Protocol (VoIP) technologies such as Skype also makes use of audio (and sometimes video) channels and could be classified as CMD using the proposed scheme.

“Anonymous messaging”, “private messaging”, “filtering”, and “quoting” all refer here to technological affordances of CMC systems. It is possible for users to engage in these behaviors without any special technical means, but when such means are available, they facilitate the behaviors, presumably making them more likely to occur. Thus, many chat systems require a user to select a nickname that is different from his or her email address, encouraging the use of pseudonyms and anonymous interaction (Danet 1998). Some Web-based discussion forums have registration procedures that do not verify users’ email addresses, encouraging users to make them up. Anonymity has been found to have important effects in online discourse, including increased self-disclosure (Kiesler et al. 1984), antisocial behavior (Donath 1999), and play with identity (Danet 1998).

Similarly, some chat systems (such as IRC and MUDs) have commands that enable users to carry on private as well as public conversations, while with other systems (such as some forms of Web chat), it is necessary to open a separate program (such as an instant messaging client) to converse privately. Along the same lines, a user can always choose to ignore messages from another user, but a number of CMC systems make this easier by providing technical mechanisms to filter out such messages (known variously as “kill files”, “gag” commands, etc.). CMC systems also differ in the extent to which they provide mechanisms to facilitate the quoting of a portion of a previous message in a response. Some email clients provide the text of the message being replied to in the new message, as a default. In others, one must copy and paste in the quoted portions manually. Severinson-Eklundh (Severinson-Eklundh & Macdonald 1994; Severinson-

Eklundh forthcoming) has observed that this can affect the extent and manner in which quoting is used.

Finally, “message format” determines the order in which messages appear, what information is appended automatically to each and how it is visually presented, and what happens when the viewing window becomes filled with messages. Most CMC systems add new messages to the bottom of a list in the order received by the system, although this is not true of blogs (which add the newest message on the top), wikis (which allow users to choose where their content will be inserted), or some experimental systems. Herring (1999) has observed that systems that post messages in the order in which they are received – which is to say most chat and discussion forums – result in disrupted turn adjacency and interleaved exchanges. The information provided in message headers (as in email) and leaders (as in chat systems) has been found to affect online self-reference and addressivity practices (Herring 1996; Werry 1996). Scrolling direction determines which messages are on the “top of the deck” and hence more likely to receive a response.

The list of medium factors in table 1 is open-ended. It is expected that some factors will be added, others further sub-divided, and others perhaps omitted as new systems are developed and researchers’ understanding of the effects of technological affordances on mediated communication deepens over time.

4.3 Situation factors

Various social and situational factors have been observed to condition variation in computer-mediated discourse (cf. Baym 1995) as in spoken discourse (cf. Hymes 1974). The set of features summarized in table 3 incorporates elements from Hymes’ SPEAKING mnemonic (see figure 1) and factors identified by Baym (1995), along with additional factors found in empirical CMD research to affect online language use. As with the medium factors, this list is not presumed to be exhaustive.

S1	Participation structure	<ul style="list-style-type: none"> • One-to-one, one-to-many, many-to-many • Public/private • Degree of anonymity/pseudonymity • Group size; number of active participants • Amount, rate, and balance of participation
S2	Participant characteristics	<ul style="list-style-type: none"> • Demographics: gender, age, occupation, etc. • Proficiency: with language/computers/CMC • Experience: with addressee/group/topic • Role/status: in “real life”; of online personae • Pre-existing sociocultural knowledge and interactional norms • Attitudes, beliefs, ideologies, and motivations
S3	Purpose	<ul style="list-style-type: none"> • Of group, e.g., professional, social, fantasy/role-playing, aesthetic, experimental • Goal of interaction, e.g., get information, negotiate consensus, develop professional/social relationships, impress/entertain others, have fun
S4	Topic or Theme	<ul style="list-style-type: none"> • Of group, e.g., politics, linguistics, feminism, soap operas, sex, science fiction, South Asian culture, medieval times, pub • Of exchanges, e.g., the war in Iraq, pro-drop languages, the project budget, gay sex, vacation plans, personal information about participants, meta-discourse about CMC
S5	Tone	<ul style="list-style-type: none"> • Serious/playful • Formal/casual • Contentious/friendly • Cooperative/sarcastic, etc.
S6	Activity	<ul style="list-style-type: none"> • E.g., debate, job announcement, information exchange, phatic exchange, problem solving, exchange of insults, joking exchange, game, theatrical performance, flirtation, virtual sex
S7	Norms	<ul style="list-style-type: none"> • Of organization

		<ul style="list-style-type: none"> • Of social appropriateness • Of language
S8	Code	<ul style="list-style-type: none"> • Language, language variety • Font/writing system

Table 3. Situation factors

“Participation structure” refers to the number of participants in the online communication situation (both actual, i.e., actively participating, and potential); the amount and rate of participation (described impressionistically or quantitatively); whether the communication is public, semi-private, or private; the extent to which interlocutors *choose* to interact anonymously/pseudonymously as opposed to in their “real life” identities¹¹ (Myers 1987); and the distribution of participation across individuals – i.e., whether participation is roughly evenly distributed, or whether some individuals or groups dominate (Herring 1993). Participation structure has implications for, among other things, politeness: public CMD tends to be less polite than private CMD (Herring 2002), and individuals who post anonymously tend to “flame” more than individuals who post in their offline identities (cf. Donath 1999).

“Participant characteristics” describe participants’ backgrounds, skills, and experiences, as well as the real life knowledge, norms, and interactional patterns they bring to bear when they engage with others online (Baym 1995). For example, participant gender has been found to affect behavior related to politeness and contentiousness within a social MUD (Cherny 1994) in two otherwise similar academic discussion lists (Herring 1996) and in a mostly-female Usenet newsgroup devoted to television soap operas as compared with norms of interaction elsewhere on Usenet (Baym 1996). Participants’ attitudes, beliefs, ideologies, and motivations relevant to their

¹¹ This value should be assigned independently of how easy or difficult the system makes sending anonymous messages or using pseudonyms. Assuming that the medium does not preclude such choices, this value encodes the extent to which users in a particular discourse sample make use of them.

online communication may also affect what they choose to communicate and how. Participants with ideological differences may be more likely to become involved in conflict discourse, as, for example, in Hodsdon-Champeon's (forthcoming) study of Usenet newsgroups on the topic of racism.

"Purpose" is potentially relevant on two levels: "Group purpose" refers in general terms to a computer-mediated group's official *raison d'être* (professional, social, etc.), while "goals of interaction" are what individual participants hope to accomplish through any given interaction; these need not, of course, be the same for any two individuals in the same interaction. Even when the same technologies are used, CMD can vary according to purpose; for example, Herring and Nix (1997) found differences in topics discussed as well as strategies for topic development in pedagogical and social IRC.

"Activities" (similar to Hymes' "genres") are discursive means of pursuing interactional goals (e.g., "flirting" as a means of developing personal relationships; "debate" as a means of impressing others with one's intellectual acumen); each activity has associated conventional linguistic practices that signal when that activity is taking place (cf. "contextualization cues", Gumperz 1982). Many studies have noted the existence of computer-mediated contextualization cues, ranging from emoticons to user IDs (Bechar-Israeli 1995; Danet et al. 1997; Heisler & Crabill 2006; Herring 2001), that help to signal "what is going on" in online interaction. Flaming, or the exchange of hostile message content, also has characteristic syntactic and semantic structures that distinguish it from other computer-mediated activity types (Spertus 1997).

"Topic" at the group level indicates, within broad parameters, what discussion content is appropriate in that context, according to the group's definition. Some CMC modes not conceived as discussion forums but rather as role-playing environments, such as adventure MUDs, may have a geographical and/or temporal "Theme" (such as a medieval village) instead of a topic. In contrast, topic at the exchange level is what participants are actually talking about in any given interaction; this may or may not be on the "official" topic of the group. Distinctions of topic are

important in analyzing topical digression, which has been claimed to be a characteristic of multi-participant text-based CMD (Herring 1999).

“Tone” refers to the manner or spirit in which discursive acts are performed (cf. Hymes’ “key”); it can be described along a number of continuous scalar dimensions, including (but not restricted to) degree of seriousness, formality, contentiousness, and cooperation. Contentious debaters on Usenet (Hodsdon-Champeon forthcoming) employ direct quoting of a discourse participant differently than do participants in friendly CMD. Emoticons similarly take on different pragmatic meanings depending on the tone of an exchange, which they may also help to establish (Huls 2006).

“Norms” refer to conventional practices within the computer-mediated environment and comprise three types. “Norms of organization” refer to formal or informal administrative protocols having to do with how a group is formed (if applicable), how new members are admitted, whether it has a leader, moderator, or other persons whose role it is to perform official functions, how messages are distributed and stored (if this is determined by social convention rather than by the system software), how participants who misbehave are punished, etc. “Norms of social appropriateness” refer to the behavioral standards that normatively apply in the computer-mediated context (cf. Hymes’ “norms of interaction”); they may be implicit or written and publicly available, for example in the form of “netiquette” guidelines (Shea 1994) or lists of Frequently Asked Questions (FAQs). Supportiveness may be expected in a women’s health newsgroup, but rudeness may be expected and approved of in the newsgroup *alt.flame*, which is devoted to flaming. “Norms of language” refer to linguistic conventions particular to a group or users; these may include abbreviations, acronyms, insider jokes, and special discourse genres (Baym 1995; Cherny 1999; Rowe forthcoming).

Finally, “code” refers to the language or language variety in which computer-mediated interactions are carried out. Although English is still the most common *language* on the Internet,

and most CMC research has been carried out on English data, this situation is changing rapidly as more non-English-speaking countries gain Internet access (Danet & Herring 2007). “Language variety” includes the dialect, and where applicable, the register of language used. The default dialect is the standard, educated, written variety of the language, although regional, social class or ethnic dialects may sometimes be used (Androutsopoulos & Ziegler 2004). Register refers here to specialized sub-languages associated with conventional social roles and contexts (such as academic discourse, psychotherapeutic discourse, teacher talk); one may also identify an unmarked register, ordinary conversation, associated with the role of the “everyday” self. Choice of linguistic code in multilingual computer-mediated groups has been observed to serve different discourse functions (Androutsopoulos & Hinnenkamp 2001; Georgakopoulou forthcoming; Paolillo 1996, forthcoming).

Relatedly, “writing system” refers to the font used and its relationship to the writing system of the language: Does the communication make use of a font (such as ASCII text) based on the Roman alphabet (e.g., for languages such as English, Spanish, and French); does it transliterate a non-roman writing system (such as those of Arabic and Greek) into Roman letters/ASCII (Berjaoui 2001; Tseliga 2007); or are special non-ASCII fonts used (such as those available for Japanese, Chinese, and Korean) to represent a non-Roman writing system? Since the introduction of the Unicode character encoding standard (see Danet & Herring 2007), it has become easier to transmit a variety of languages in their native scripts via the Internet, but transliteration into roman letters persists in some contexts, and script choice may serve different pragmatic functions (e.g., Tseliga 2007).

Although in principle the eight situation dimensions in table 2 are independent of one another, in practice, they tend to combine in predictable ways. This is easiest to see when the classification scheme is applied to familiar CMC modes. For example, discourse in Internet Relay Chat typically is many-to-many, has a high degree of anonymity (participants use pseudonyms), is social in function and non-serious in tone, contains a high incidence of flirting and phatic (empty,

social) exchanges, and appears to be engaged in most often by young people between the ages of 18 and 25 (Danet et al. 1997; Reid 1991; Werry 1996). In contrast, discourse in an academic discussion list is more likely to serve professional purposes, have a serious tone, contain debates and job announcements, and be engaged in by older, professionally established users (Grüber 1998, 2000; Herring 1992, 1996; Hert 1997). Furthermore, medium factors may correlate with situation factors; all other things being equal, for example, synchronous CMD is more likely to be informal in register and playful in tone than is asynchronous CMD (Herring 2001).

However, it is important to note that there are also circumstances under which these associations do not hold. The classification scheme presented above, because it does not presume any *necessary* relationships among features of situational context or between medium and situation, allows unpredictable and unconventional associations to emerge as easily as more typical ones. This is illustrated in the following section.

5. Sample classification

While it is beyond the scope of this article to test the proposed classification scheme formally, a brief illustration of its application to two samples of CMD may provide a glimpse of the utility of the scheme. One sample is from a well-known, popular source, and the other from a closed-access, privately-developed system; both have been analyzed by the author in separate studies, albeit not from a classification perspective.¹²

Both samples are exemplars of the sociotechnical mode “weblogs” (blogs), broadly construed. Blogs have been characterized as a genre of CMC (Herring et al. 2004; Miller & Shepherd 2004), although subtypes such as diary and filter blog have also been identified that manifest distinct patterns of linguistic usage (Herring & Paolillo 2006). In the comparison

¹² The *LiveJournal* data were collected as part of the project reported in Herring et al. (2007), and a preliminary analysis of the *Quest Atlantis* blog data is reported in Herring, de Siqueira, Stuckey & Kouper (in review). *Language@Internet* 1/2007 (<http://www.languageatinternet.de>, urn:nbn:de:0009-7-7611, ISSN 1860-2029)

described below, however, it is not sufficient to distinguish subtypes, since one sample is relatively novel, and a single instance cannot form a type.

The first sample is from the popular blog-hosting service *LiveJournal.com*, which claims to have hosted over 11.9 million blogs since its inception in 1999. The second sample is from *Quest Atlantis*, a game-like online learning environment for children 9-12 years old that was developed in 2002 by researchers at the author's institution (Barab et al. 2005), and that has been used by several thousand children to date, mostly in the United States, Australia, and Singapore, under the supervision of their classroom teachers. *Quest Atlantis* (QA) includes blogs as one of several types of CMC available to its young users. Specifically, our QA sample comes from a blog maintained by a fictional Atlantian girl, Alim (in reality, an adult female QA researcher), who posts entries on the theme of "personal agency" for children on Earth; the children post comments in response.

In order to make our samples as comparable as possible, let us consider the LJ of a young, English-speaking woman. Moreover, although both sources make available data extending over a period of more than two years, let us further delimit each sample to two months of continuous activity in spring 2006. The exact time and size of the samples are not important for the purpose of this illustration, but a multi-message sample is necessary in order to obtain a sense for how discourse takes place typically, over time.

Not surprisingly, since both are known by the genre label "blog", these two samples share many medium features. These include asynchronicity (M1); 1-way message transmission (M2); persistence of messages in archives linked from the sidebar of the blog (M3); Web-based delivery and a tendency for messages to be text only (M5); and the display of blog entries in reverse chronological sequence with a "comment" option below each entry (M10). These might be considered definitional characteristics of the blog genre (see also Herring et al. 2004).

However, the two samples have few situation variables in common, aside from a one-to-many participation structure and imbalanced participation¹³ (S1), which are characteristic of blog discourse in general (Herring et al. 2004). Holding blog author gender (S2) and use of the English language (S8) constant does not result in any other associated similarities between the two samples.

In contrast, differences can be observed along both the medium and the situation dimensions. Whereas LJ allows anyone to create a blog from a made-up name (as our sample LJ blogger has done), anonymity is impossible in the QA blogs, since all users must register through their classroom teachers (M6). LJs are publicly available on the Web unless designated as “friends only” (our sample is not so designated), whereas QA activity is closed to the public (M7). There are also differences in message format (M10) – the LJ interface is more sophisticated, providing users with more options (such as “friends” links and a “search” feature) and greater social translucence (Erickson et al. 1999), such as an indication of the number of comments that have been posted after each blog entry.

The number of differences in situation between the two samples is also great. Group size, construed as the potential audience of each blog, varies widely as a consequence of the public/private nature of each blog; rate of participation is also slower on the QA blog, and posting rights are asymmetrical (S1) – only “Alim” can post entries. In the LJ, only the blog owner can post in her own blog, but commenters all have their own blogs, so everyone has a chance to both post and comment. Age, roles, previous experience, and the relationships among participants also differ between the two samples (S2), as does the purpose of each blog (S3), its topic/theme (S5), the tone of messages and comments (S6), and the norms of interaction and norms of language use in LJ versus QA (S7).

¹³ Blog owners post more and longer messages than do visitors to the blog, who typically may only post comments on the owner's entries.

The LJ blogger is an experienced, adult Internet user who posts messages about her day-to-day life to friends and strangers in a tone that aims for cleverness and sophistication, and where the norms of interaction include profanity and sexual references. In these respects, the LJ sample is typical of many LJ blogs (cf. Kendall 2005). The considerable contrast between these two samples reflects QA's young, inexperienced target audience and its educational context, which is closely moderated by adults, and which assigns asymmetrical posting rights to adults and children. These are not prototypical blog features, although the QA blogs recall other uses of CMC in primary education (Robertson, Good & Pain 1998).

Clearly, simply classifying these samples as being of the blog mode or genre, while it would capture more-or-less predictable associations for *LiveJournal*, would miss much about the QA data that is interesting and important. Moreover, the LJ data also exhibit characteristic properties that differentiate them from the blog prototype (cf. Herring et al. 2004), such as the "friends only" audience designation feature and "mood" indicators for entries. A faceted classification approach is thus revealing for LJ blogs as well, and more generally, is essential (in some form) for characterizing different blog subtypes.

6. Conclusions

As the Internet expands, it continues to spawn new varieties of discourse that call out for analysis and classification. This article has proposed, argued for, and briefly illustrated the utility of a faceted classification scheme for computer-mediated discourse. This scheme classifies discourse samples in terms of clusters of variable dimensions, thereby preserving their complexity (including overlap across samples) and allowing for focused comparisons within and across samples.

The faceted scheme is intended to complement existing mode-based classification of CMD. Mode classification is especially useful for identifying and invoking prototypical associations of CMD data of a type that is generally known, such as email, discussion lists, and

IRC; it also captures cultural information that cannot be predicted solely from the component dimensions of the scheme. However, mode classification is less useful for proprietary or novel examples of online discourse, such as the Quest Atlantis blogs or the quasi-synchronous “Babble” chat system developed by Erickson et al. (1999), which do not evoke prototypical associations except in the minds of users who happen to know the systems. Faceted classification is more useful for characterizing CMD in such cases.

At the same time, the classification scheme presented here has several limitations. First, it can seem verbose (a “list” of terms) and difficult to condense due to its relatively non-hierarchical (“flat”) structure. Selective classification, following the analytico-synthetic principle of Ranganathan (1933), in which only the most important features of a data set (as determined by the goals of the research) are selected for characterization, is recommended to help address this problem.

A second limitation is that the scheme is based primarily on research findings for textual CMC. It is important, but ultimately not sufficient, to note that multimedia CMC makes use of multiple channels of communication. Mobile and voice-over-IP communication raise additional classificatory challenges. What are the criteria for identifying types of multiplayer online game discourse, for example? What are the relevant dimensions that condition variation in video- and audio-mediated communication? What about in CMD where participants can speak, text chat, and manipulate a common interface (such as a whiteboard) at the same time? It will be essential to address these challenges in future CMC classification research.

A more general limitation is that the scheme is not in itself a contribution to a theory of genre, but is rather a preliminary aggregation of factors that will have to find a place in a theory of CMD genres. Theoretical questions remain to be addressed concerning the organization and relationships among the features of the scheme. Conversely, it is conceivable that empirical investigation of feature co-occurrence patterns based on this descriptive scheme could lead to the

identification of a smaller set of CMC prototypes. If so, these could be compared with genres already posited for Internet communication (cf. Giltrow & Stein in preparation), lending them an empirical underpinning. Investigation of this possibility and theoretical development of the scheme itself are desiderata for future research.

Finally, Hymes cautions that “an ‘etic’ account, however useful as a preliminary grid and input to an emic (structural) account, or as a framework for comparing different emic accounts, lacks the emic account’s validity” (1974: 11). Simple descriptive classification should be supplemented by ethnographic observation of online discourse communities over time, and should ideally be validated by members of those communities, in order to provide the richest possible context for the analysis of computer-mediated discourse.

References

- Anderson, Jeffery F., Fred K. Beard, & Joseph B. Walther (forthcoming). The local management of computer-mediated conversation. In Herring, Susan C. (ed.).
- Androutsopoulos, Jannis (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10(4): 419-438.
- Androutsopoulos, Jannis & Volker Hinnenkamp (2001). Code-switching in der bilingualen Chat-Kommunikation: ein explorativer Blick auf #hellas und #turks. In Beisswenger, Michael (ed.). 367-401.
- Androutsopoulos, Jannis & Evelyn Ziegler (2004). Exploring language variation on the Internet: Regional speech in a chat community. In Gunnarsson, Britt-Louise, Lena Bergström, Gerd Eklund, Staffan Fridell, Lise H. Hansen, Angela Karstadt et al. (eds.) *Language Variation in Europe: Papers from the Second International Conference on Language Variation in Europe, ICLaVE 2*. Uppsala: Uppsala University. 99-111.
- Anis, Jacques (2007). Neography: Unconventional spelling in French SMS text messages. In Danet, Brenda & Susan C. Herring (eds.) *The multilingual Internet: Language, culture, and communication online*. New York: Oxford University Press.
- Barab, Sasha A., Michael Thomas, Tyler Dodge, Robert Carteaux, & Hakan Tuzun (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development* 53(1): 86-107.
- Baron, Naomi (forthcoming). Discourse structures in instant messaging: The case of utterance breaks. In Herring, Susan C. (ed.).
- Baym, Nancy (1995). The emergence of community in computer-mediated communication. In Jones, Steven G. (ed.). 138-163.
- Baym, Nancy (1996). Agreements and disagreements in a computer-mediated discussion. *Research on Language and Social Interaction* 29(4): 315-345.

- Bechar-Israeli, Haya (1995). From (Bonehead) to (cLoNehEAd): Nicknames, play and identity on Internet relay chat. *Journal of Computer-Mediated Communication* 1(2).
<http://jcmc.indiana.edu/vol1/issue2/bechar.html>.
- Beisswenger, Michael (ed.) (2001). *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld*. Stuttgart: Ibidem.
- Berjaoui, Nasser (2001). Aspects of the Moroccan Arabic orthography with preliminary insights from Moroccan computer-mediated communication. In Beisswenger, Michael (ed.). 431-465.
- Biber, Douglas (1988). *Variation in speech and writing*. Cambridge, UK: Cambridge University Press.
- Broughton, Vanda & Heather Lane (2000). Classification schemes revisited: Applications to Web indexing and searching. *Journal of Internet Cataloguing* 2(3/4): 143-155.
- Chafe, Wallace L. & Jane Danielewicz (1987). Properties of spoken and written language. In Horowitz, Rosalind & S. Jay Samuels (eds.) *Comprehending oral and written language*. New York: Academic. 83-113.
- Cherny, Lynn (1994). Gender differences in text-based virtual reality. In Bucholtz, Mary, Anita Liang, & Laurel Sutton (eds.) *Cultural Performances: Proceedings of the Third Berkeley Women and Language Conference*. Berkeley: Berkeley Women and Language Group.
- Cherny, Lynn (1999). *Conversation and community: Chat in a virtual world*. Stanford, CA: Center for the Study of Language and Information.
- Chou, Candace C. (1999). From simple chat to virtual reality: Formative evaluation for synchronous communication systems to online learning. *WebNet 1999*. 225-230. Available at <http://www2.hawaii.edu/~cchou/ppdla99/index.htm>.
- Condon, Sherri L. & Claude G. Cech (1996). Discourse management strategies in face-to-face and computer-mediated decision making interactions. *Electronic Journal of Communication* 6(3).
<http://www.cios.org/www/ejc/v6n396.htm>.

- Condon, Sherri L. & Claude G. Cech (2001). Profiling turns in interaction. *Proceedings of the 34th Annual Conference of the Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.
- Crystal, David (2001). *Language and the Internet*. Cambridge, UK: Cambridge University Press.
- Danet, Brenda (1998). Text as mask: Gender, play and performance on the Internet. In Jones, Steven G. (ed.). 129-158.
- Danet, Brenda & Herring, Susan C. (2007). Multilingualism on the Internet. In Hollinger, Marlis & Anne Pauwels (eds.) *Language and communication: Diversity and change. Handbook of applied linguistics, vol. IX*. Berlin: Mouton de Gruyter.
- Danet, Brenda, Lucia Ruedenberg & Yehudit Rosenbaum-Tamari (1997). "Hmmm ... Where's that smoke coming from?" Writing, play and performance on Internet Relay Chat. In Rafaeli, Sheizaf, Fay Sudweeks & Margaret McLaughlin (eds.) *Network and netplay: Virtual groups on the Internet*. Cambridge, MA: AAI/MIT Press. 41-76.
- Donath, Judith (1999). Identity and deception in the virtual community. In Smith, Marc A. & Peter Kollock (eds.) *Communities in cyberspace*. London: Routledge. 29-59.
- Dooley, Robert A. & Stephen H. Levinsohn (2001). *Analyzing discourse: A manual of basic concepts*. Dallas: SIL International.
- Erickson, Thomas, David N. Smith, Wendy A. Kellogg, Mark R. Laff, John T. Richards, & Erin Bradner (1999). Socially translucent systems: Social proxies, persistent conversation, and the design of 'Babble'. In *Human Factors in Computing Systems: Proceedings of CHI '99*. ACM Press.
- Ferrara, Kathleen, Hans Brunner & Greg Whitemore (1991). Interactive written discourse as an emergent register. *Written Communication* 8(1): 8-34.
- Georgakopoulou, Alexandra (forthcoming). 'On for drinkies?': E-mail cues of participant alignments. In Herring, Susan C. (ed.).

- Giltrow, Janet & Dieter Stein (eds.) (in preparation). *Theories of genre and their application to Internet communication*.
- Grüber, Helmut (1998). Computer-mediated communication and scholarly discourse: Topic initiation and thematic development. *Pragmatics* 8(1): 21-46.
- Grüber, Helmut (2000). Scholarly email discussion postings: A single new genre of academic communication? In Pemberton, Lyn & Simon Shurville (eds.) *Words on the Web: Computer-mediated communication*. Exeter: Intellect. 36-43.
- Gumperz, John J. (1982). Contextualization conventions. *Discourse strategies*. Cambridge, UK: Cambridge University Press. 130-152.
- Heisler, Jennifer & Scott Crabill (2006). Who are “stinkybug” and “packerfan4”? Email pseudonyms and participants' perceptions of demography, productivity, and personality. *Journal of Computer-Mediated Communication* 12(1), article 6.
<http://jcmc.indiana.edu/vol12/issue1/heisler.html>.
- Herring, Susan C. (1992). *Gender and participation in computer-mediated linguistic discourse*. Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics. Document no. ED345552.
- Herring, Susan C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication* 3(2). <http://ella.slis.indiana.edu/~herring/ejc.txt>.
- Herring, Susan C. (ed.) (1996). *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins.
- Herring, Susan C. (1996). Two variants of an electronic message schema. In Herring, Susan C. (ed.). 81-106.
- Herring, Susan C. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication* 4(4). <http://jcmc.indiana.edu/vol4/issue4/herring.html>.
- Herring, Susan C. (2001). Computer-mediated discourse. In Tannen, Deborah, Deborah Schiffrin & Heidi Hamilton (eds.) *Handbook of discourse analysis*. Oxford: Blackwell. 612-634.
- Language@Internet* 1/2007 (<http://www.languageatinternet.de>, urn:nbn:de:0009-7-7611, ISSN 1860-2029)

- Herring, Susan C. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology* 36: 109-168.
- Herring, Susan C. (2004a). Computer-mediated discourse analysis: An approach to researching online behavior. In Barab, Sasha A., Rob Kling & James H. Gray (eds.) *Designing for virtual communities in the service of learning*. New York: Cambridge University Press. 338-376.
- Herring, Susan C. (2004b). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society* 6(1): 26-36.
- Herring, Susan C. (ed.) (forthcoming). *Computer-mediated conversation*. Cresskill, NJ: Hampton Press.
- Herring, Susan C., Amaury de Siqueira, Bronwyn Stuckey & Inna Kouper (in review). *Educational blogs for children: From conversation to community*.
- Herring, Susan C., Anna Martinson & Rebecca Scheckler (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, CA: IEEE Press.
- Herring, Susan C. & Carole G. Nix (1997). *Is 'serious chat' an oxymoron? Academic vs. social uses of Internet Relay Chat*. Paper presented at the American Association of Applied Linguistics, Orlando, FL, March 11.
- Herring, Susan C. & John C. Paolillo (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics* 10(4): 439-459.
- Herring, Susan C., John C. Paolillo, Irene Ramos-Vielba, Inna Kouper, Elijah Wright, Sharon Stoerger, Lois Ann Scheidt & Benjamin Clark (2007). Language networks on LiveJournal. *Proceedings of the Fortieth Hawai'i International Conference on System Sciences*. Los Alamitos, CA: IEEE Press.

- Herring, Susan C., Lois Ann Scheidt, Sabrina Bonus & Elijah Wright (2004). *Bridging the gap: A genre analysis of weblogs. Proceedings of the 37th Hawai'i International Conference on System Sciences*. Los Alamitos: IEEE Press.
- Hert, Philippe (1997). Social dynamics of an on-line scholarly debate. *The Information Society* 13: 329-360.
- Hodsdon-Champeon, Connie (forthcoming). Conversations within conversations: Intertextuality in racially antagonistic dialogue on Usenet. In Herring, Susan C. (ed.).
- Huls, Erica (2006). *The communicative functions of emoticons in computer-mediated communication*. Unpublished manuscript.
- Hymes, Dell (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Jones, Steven G. (ed.) *Cybersociety: Computer-mediated communication and community*. Thousand Oaks, CA: Sage.
- Kendall, Lori (2005). Diary of a networked individual: System design's effects on online relationships. In Consalvo, Mia (ed.) *Internet research annual*. New York: Peter Lang. 41-50.
- Kiesler, Sara, Jane Siegel & Timothy W. McGuire (1984). Social psychological aspects of computer-mediated communication. *American Psychologist* 39: 1123-1134.
- Ko, Kwang-Kyu (1996). Structural characteristics of computer-mediated language: A comparative analysis of InterChange discourse. *Electronic Journal of Communication* 6(3).
<http://www.cios.org/www/ejc/v6n396.htm>.
- Longacre, Robert (1996). Typology and salience. *The grammar of discourse*, 2nd edition. New York: Plenum Press. 7-31.
- Mangueneau, Dominique (2002). Analysis of an academic genre. *Discourse Studies* 4 (3): 319-342.

- Markus, M. Lynne (1994). Finding a happy medium: Explaining the negative effects of electronic communication on social life at work. *ACM Transactions on Information Systems* 12(2): 119-149.
- Maynor, Natalie (1994). The language of electronic mail: Written speech? In Montgomery, Michael & Greta D. Little (eds.) *Centennial usage studies*. Publications of the American Dialect Society Series. Tuscaloosa : Published for the Society by the University of Alabama Press.
- Miller, Caroline R. & Dawn Shepherd (2004). Blogging as social action: A genre analysis of the weblog. In Gurak, Laura J., Smiljana Antonijevic, Laurie Johnson, Clancy Ratliff & Jessica Reyman (eds.) *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. Minneapolis: University of Minnesota. Available at http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html
- Murray, Denise E. (1988). The context of oral and written language: A framework for mode and medium switching. *Language in Society* 17: 351-373.
- Murray, Denise E. (1990). CmC. *English Today* 23: 42-46.
- Myers, David (1987). 'Anonymity is part of the magic': Individual manipulation of computer-mediated communication environments. *Qualitative Sociology* 19(3): 251-266.
- Nevins, M. Eleanor (2004). Learning to listen: Confronting two meanings of language loss in the contemporary White Mountain Apache speech community. *Journal of Linguistic Anthropology* 14(2): 269.
- Paolillo, John C. (1996). Language choice on soc.culture.punjab. *Electronic Journal of Communication* 6(3). <http://www.cios.org/www/ejc/v6n396.htm>.
- Paolillo, John C. (forthcoming). Conversational codeswitching on Usenet and Internet Relay Chat. In Herring, Susan C. (ed.).

Prieto-Diaz, Ruben (1991). Implementing faceted classification for software reuse.

Communications of the ACM 34(5): 88-97.

Ranganathan, S. R. (1933). *Colon classification*. (1st edition.) Madras: Madras Library Association.

Reid, Elizabeth M. (1991). *Electropolis: Communication and community on Internet Relay Chat*.

Unpublished senior honours thesis, University of Melbourne, Australia. Available at <http://www.aluluei.com/>.

Rice, Ron & Urs E. Gattiker (2000). New media and organizational structuring. In Jablin, Fredric & Linda L. Putnam (eds.) *The new handbook of organizational communication*. Thousand Oaks, CA: Sage. 544-581.

Robertson, Judy, Judith Good & Helen Pain (1998). BetterBlether: The design and evaluation of a discussion tool for education. *International Journal of Artificial Intelligence in Education* 9: 219-236.

Rowe, Charley (forthcoming). Genesis and evolution of an e-mail-driven sibling code. In Herring, Susan C. (ed.).

Severinson Eklundh, Kersten (1986). *Dialogue Processes in Computer-Mediated Communication. A Study of Letters in the COM System*. Linköping Studies in Arts and Science 6. Department of Communication Studies, Linköping University.

Severinson Eklundh, Kersten (forthcoming). To quote or not to quote: Setting the context for computer-mediated dialogues. In Herring, Susan C. (ed.).

Severinson Eklundh, Kersten & Clare Macdonald (1994). The use of quoting to preserve context in electronic mail dialogues. *IEEE Transactions on Professional Communication* 37(4): 197-202.

Shea, Virginia (1994). *Netiquette*. San Francisco: Albion.

Spertus, Ellen (1997). Smokey: Automatic recognition of hostile messages. *Innovative Applications of Artificial Intelligence (IAAI) '97*.

- Swales, John (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tseliga, Theodora (2007). "It's all *Greeklisch* to me!": Linguistic and sociocultural perspectives on Roman-alphabetized Greek in asynchronous computer-mediated communication. In Danet, Brenda & Susan C. Herring (eds.).
- Tudhope, Douglas, Ceri Binding, Dorothee Blocks & Daniel Cunliffe (2002). Representation and retrieval in faceted systems. In López-Huertas, María J. & Francisco J. Muñoz-Fernández (eds.) *Advances in knowledge organization 8*: 191-197. Würzburg: Ergon.
- Vickery, Brian C. (1960). *Faceted classification: A guide to construction and use of special schemes*. London: Aslib.
- Virtanen, Tuija (1992). Issues of text typology: Narrative – a 'basic' type of text? *Text* 12(2): 293-310.
- Werry, Christopher C. (1996). Linguistic and interactional features of Internet Relay Chat. In Herring, Susan C. (ed.). 47-63.
- Yates, Simeon J. (1996). Oral and written linguistic aspects of computer conferencing. In Herring, Susan C. (ed.). 29-46.