

Running head: IM AND GENDER

SEE YOU ONLINE:

Gender Issues in College Student Use of Instant Messaging

Naomi S. Baron

Department of Language and Foreign Studies

American University

Abstract

Instant Messaging (IM) is becoming a mainstay for online one-to-one communication. Although IM is popularly described as a written version of informal speech, little empirical investigation of the linguistic nature of IM exists. Moreover, while gender issues are being addressed for one-to-many forms of computer mediated communication, we have no comparable studies of IM. This article offers a linguistic profile of American college student IM conversations. In addition to analyzing conversational scaffolding and lexical issues, the article identifies gender divergences in IM usage. Some differences reflect commonly reported functional gender distinctions in face-to-face spoken conversation; other differences indicate gender-based attitudes towards the importance of language standards in speech and writing.

See You Online:

Gender Issues in College Student Use of Instant Messaging¹

“See you online,” said one teenaged girl to another, as they left school for the day, each headed for her respective home – and connection to the Internet. By “online,” the speaker meant Instant Messaging (IM), through which a rapidly growing number of American teenagers communicate with one another (Lenhart, Rainie, & Lewis, 2001). Much has been written in the popular press (e.g., Helderman, 2003; Lee, 2002) about teenage use of IM. While there are a handful of empirically-based studies of teenage IM (e.g., Boneva, Quinn, Kraut, Kiesler, & Shklovski, in press; Grinter & Palen, 2002; Randall, 2002; Schiano et al., 2002), there is almost no research on the linguistic characteristics of these conversations (Jacobs, 2003 is an exception). Even less is known about IM behavior as these users move into college and adulthood. Jones (2002) notes the popularity of IM on American college campuses, though outside of Hård af Segerstad’s work (2002) on a form of IM used in a specialized Swedish setting, we lack empirical studies of the linguistic nature of IM conversations constructed by perhaps one of the greatest populations of its users, college students.

Instant messaging is one form of the larger online phenomenon of computer mediated communication (CMC). The term CMC refers to a cluster of interpersonal communication systems used for conveying written text, generally over the Internet. The two major parameters across which types of CMC most significantly differ are first, whether they are synchronous or asynchronous (i.e., whether or not transmission is essentially instantaneous and interlocutors are assumed to be physically present to read messages and respond to them) and second, whether the communication is one-to-one (i.e.,

between two people) or one-to-many (i.e., one person's message is broadcast to multiple potential interlocutors). Schematically, the four classes of CMC are as follows:

	Synchronous	Asynchronous
one-to-one	IM	email, texting on mobile phones
one-to-many	Chat, MUDs, MOOs, ² computer conferencing	listservs, newsgroups

In one-to-one forms of CMC (with the partial exception of email), the interlocutors generally know one another, while with one-to-many forums, they often do not.

Linguists and other social scientists have been studying CMC for well over a decade (e.g., *Journal of Computer-Mediated Communication*; Baron, 1984, 1998, 2003; Collott & Belmore, 1996; Crystal, 2001; Ferrara, Brunner, & Whittemore, 1991; Herring, 1996, 2002; Maynor, 1994; Yates, 1996). However, we need to be aware that each type of CMC has its own usage conditions and therefore, each needs to be analyzed in its own right. These usage conditions may, in turn, influence the character of language produced in that medium (e.g., formal versus informal, collaborative versus aggressive, verbose versus terse, edited versus scattershot, informative versus whimsical).

One important question about the language we find in CMC is the extent to which male and female users of the medium manifest different linguistic patterns. Within the field of sociolinguistics, the topic of gender differences in language has a long history (for reviews of the literature, see Cameron, 1998; Eckert, 1989; Holmes & Meyerhoff, 2003). Most studies are based on spoken language behavior, though a small body of research has examined the possible effects of gender on written style. Internet researchers have also been exploring gender-based correlates of online behavior. Nearly all of the work to date

has drawn upon one-to-many data sources (e.g., Chat, listservs, computer conferencing). We know very little about the kinds of language that male versus female interlocutors create in one-to-one online venues such as email (though see Boneva & Kraut, 2002; Thomson, Murachver, & Green, 2001) or IM.

The empirical goals of this article are two-fold. The first is to develop a linguistic profile of IM conversations used by contemporary American undergraduate students. The second is to determine whether the linguistic patterns in these conversations reveal gender-based distinctions.

This study begins by outlining a linguistic framework for analyzing IM as a form of CMC. The following section summarizes relevant literature on gender and speech, gender and writing, and gender and CMC. The bulk of the article is devoted to an empirical analysis of an IM corpus collected in Spring 2003, including discussion of linguistic domains that reveal gender differences. The article concludes by examining the extent to which IM conversational behavior reflects or diverges from gender-based language patterns previously reported in spoken and traditional written venues.

Instant Messaging as a Form of Computer Mediated Communication

Technologically, IM is a synchronous form of one-to-one CMC. Although only two people are engaged in each conversation, users may conduct multiple IM conversations simultaneously or IM while using their computers for other functions (e.g., word processing, Web searches, commercial transactions) at the same time, not to mention engaging in simultaneous off-line activities (e.g., talking, watching television, eating). Therefore, it is common for interlocutors not to have one another's undivided attention, as they might better command in face-to-face conversations. The current version (Spring

2004) of America Online Instant Messenger (AIM), which is the IM system used by the majority of American college students, informs users when someone is in the process of typing a message to them. However, the potential recipient does not see any of the actual message until it has been sent.

As a form of conversation between two individuals, IM exchanges can, in principle, be viewed in terms of the same kinds of linguistic variables as used in analyzing spoken face-to-face discourse between two interlocutors (see Sachs, Schegloff, & Jefferson, 1974; Schegloff & Sachs, 1973; Schiffrin, Tannen, & Hamilton, 2001 for discussion of the analysis of spoken language discourse). For our purposes, we divide such variables into four groups: individual turns, combining turns into conversational sequences, openings and closings, and conversation management.

Individual turns. The simplest way of looking at IM conversations is at the level of individual turns. We can ask such quantifiable questions as how long these turns are (in words), how many turns there are per minute, and how often turns are made up of a single word rather than multiple words. In examining the words (and word-like symbols) that make up individual turns, we can consider such issues as how many words are abbreviated or contracted, how many emoticons are used, what kind of punctuation appears, or how frequently we find spelling errors or self-corrections.

Combining turns into conversational sequences. The next step up in discourse complexity is to consider how turns are combined into sequences and conversations. For example, we need to look at sequences of turns used by a sender (i.e., without interruption from his or her interlocutor), at how users break up single sentences into multiple turns, and

at how the larger conversation is structured (e.g., how long are IM conversations – with respect to both total number of turns and clock time).

Openings and closings. An important subset of discourse construction is how interlocutors open and close their conversations. Are there special formulae used within a dyad or a larger discourse group? How many turns (and/or how much time on the clock) does it take to initiate or terminate an IM exchange?

Conversation management. The most complex level of IM discourse analysis is that of conversation management. How do the interlocutors structure the body of their conversation? Some of the variables at issue are the same as in spoken conversation: How do you take and hold the floor? What constitutes a felicitous reply to an interlocutor's utterance? When may you interrupt a chain of transmissions (or a chain of thought)? In addition, IM technology introduces the possibility of message transmissions overlapping, with the result that a new conversational thread may be introduced before an earlier topic is concluded. How do interlocutors sort out multiple simultaneous conversation threads?

Beyond these four discourse domains lie various meta-issues. One of these is the degree of attention to the conversation at hand. Many people sometimes use IM asynchronously, wandering away from the computer for minutes or even hours at a time (i.e., while still logged on), conducting multiple simultaneous conversations, and even editing an IM turn they have typed before actually sending it (Nardi, Whittaker, & Bradner, 2000; Gloria Jacobs, personal communication, October 2004). Other meta-issues affecting IM discourse originate in variables going beyond the scope of the conversation itself. Among these variables are age of the interlocutors, context of usage (e.g., personal versus business), conditions of usage (e.g., private messaging sent from an office computer;

teenagers with parents looking over their shoulders), local or national culture, extent of experience with the medium, and, of course, gender.

Gender and Language

Linguists have long recognized that a person's gender may affect his or her linguistic productions. At the most obvious level, some languages of the world restrict particular lexical, phonological, or grammatical usage patterns to males or females. In Australia, for example, there is a group of aboriginal women who use a sign language that males are forbidden to learn (Kendon, 1980). Other gender differences are the result of more subtle acculturation. For example, it is commonly reported that females tend to use more politeness indicators than males (e.g., Coates, 1993), while men more frequently interrupt women than vice versa (e.g., Tannen, 1994). Some of these differences appear not just in the United States or the west more generally, but cross-culturally (Chambers, 1992; Holmes, 1993). If we hope to understand potential gender-based differences in instant messaging conversations, it behooves us to understand how gender distinctions work in off-line language.

Gender analysis of IM is further complicated by the fact that IM (like other forms of CMC) is arguably based on both spoken and written premises. That is, while the messages themselves are physically written and bear some characteristics of written language, they also have important resemblances to speech (Baron, 2000; Crystal, 2001). We therefore need to look not only at the literature on gender and speech but also at studies on gender and writing. In addition, the growing body of research on gender issues in one-to-many CMC platforms may help illuminate gender analysis of a one-to-one CMC platform such as IM.

Speech and Gender

The majority of studies concerning language and gender have been based upon the analysis of spoken language, drawing upon direct observation, interviews, or transcriptions appearing in large-scale corpora. Two of the themes most commonly addressed in these studies have been first, the use of social versus more neutrally informative speech, and second, the extent to which speech adheres to normative standards.

Social versus informative. Sociolinguists (e.g., Cameron, 1998; Coates, 1993; Eckert & McConnell-Ginet, 2003; Holmes, 1993; Romaine, 2003; Tannen, 1994) have commented on the tendency of women (largely, though not exclusively in the west) to use conversation predominantly as a tool for facilitating social interaction, while their male counterparts are more prone to use conversation for conveying information. In Holmes' words (1995, p. 2), while women "use language to establish, nurture and develop personal relationships," men's use of conversation is more typically "a means to an end". These two social constructs (social versus informative) derive from observations of concrete linguistic features in which the speech of at least many women diverges from that of at least many men. For example, women tend to use more affective markers (e.g., "I know how you feel"), more diminutives (e.g., "little bitty insect"), more hedge words (e.g., *perhaps*, *sort of*), more politeness markers (e.g., "I hate to bother you"), and more tag questions (e.g., "We're leaving at 8:00 pm, aren't we?") than do men. Men, on the other hand, are likely to use more referential language (e.g., "The stock market took a nosedive today"), more profanity, and fewer first person pronouns than are women.

Another way of looking at the social function of language is to ask how much talking takes place. Do women (as commonly assumed) talk more than men? Multiple

findings reported by Holmes (1993) suggest that males rather than females tend to dominate public conversations among mixed-sex participants. While there is now a growing literature on same-sex dyadic conversation between friends (e.g., Coates, 1998), there seems to be a paucity of studies specifically comparing same-sex male and female verbosity, particularly among adolescents or young adults.

Standard versus substandard usage. A second common finding in the gender and language literature has been that on average, women's speech reflects standard phonological, lexical, and grammatical patterns more than men's does (e.g., Chambers, 1992; Holmes, 1993; James, 1996; Labov, 1991). More detailed analysis has revealed this pattern to be especially strong among lower-middle class females versus males (see Labov, 1991 for a summary of the evidence). A variety of explanations have been offered for the gender discrepancies: that women are socialized to speak more "correctly"; that women are interested in social mobility; that at least among children and young adults, there is a biological basis for women's higher verbal abilities (Chambers, 1992); or that since women tend to have less social power than men, language is one of the few domains in which women can exercise social superiority. Without entering into this larger debate, we can note that especially in the modern west, women's social roles in the family and in the public sphere have made acquisition and use of standard language patterns personally advantageous. On the one hand, since women do the majority of the child rearing, they can model standard language usage for their progeny. On the other hand, in much of the 20th century, when women's professional choices were largely circumscribed, the positions that were broadly open to women (e.g., teacher, secretary, airline stewardess) required their incumbents to be well-spoken.

Writing and Gender

There are a small number of studies pertaining to gender differences in written language. Some of the data analyzed are historical in nature (relying heavily upon personal letters), while other data derive from large-scale written corpora or experimental essay composition tasks. In most empirical analyses, the same two themes we looked at regarding gender and speech have been judged to be significant, namely the extent to which texts function socially or referentially, and the extent to which writers adhere to normative standards. However, rather than organize our synopsis of the literature in terms of these two categories, we will first look at findings by genre (specifically, personal letters) and research agenda (identifying gender via textual analysis), and then consider national measurements of children's writing skills in the United States. We will then summarize the findings in terms of the social versus informational and standard versus non-standard dimensions.

Personal letters. Biber and his colleagues (Biber, 1988; Biber, Conrad, & Reppen, 1998; Biber & Finegan, 1997) have studied speech and writing by using corpus-based data drawn from a variety of venues, enabling them to analyze register variation across language use contexts. The five major dimensions of analysis Biber identified were (1) involved versus informational production, (2) narrative versus non-narrative discourse, (3) elaborated versus situation-dependent reference, (4) overt expression of argumentation, and (5) impersonal versus non-personal style. Each dimension represents a cluster of linguistic features. For example, "involved versus informational production" measures such variables as use of present tense verbs, appearance of first- and second-person pronouns, use of contractions, and use of "private" verbs such as *think* or *feel*. Note that the "involved versus

informational” dimension roughly parallels the “social versus informative” dichotomy that has commonly been used to distinguish (at least statistically) between female versus male speech patterns.

Biber et al. (1998) applied this linguistic framework to a corpus of personal letters written by both men and women over the course of four centuries (from the 17th through the 20th), particularly with regard to the dimension “involved versus informational.” Included in their sample were letters written by women to women (FF), by women to men (FM), by men to women (MF), and by men to men (MM). While the number of letters analyzed was small (especially in the FF and FM samples), Biber et al. report some interesting gender differences, as well as changes in gender patterns over time. In both the 17th and the 20th centuries, personal letters written by women (i.e., FF and FM) showed a higher index of “involved” language than did letters written by men (i.e., MF and MM).

In a related study, Palander-Collin (1999) examined personal letters written by men and women in the 17th century. Focusing on the phrase “I think,” which combines two elements of Biber’s dimension of involvement (i.e., the first person pronoun *I* and the private verb *think*), Palander-Collin confirmed Biber et al.’s findings that personal letters written by women showed higher levels of interpersonal involvement than did letters written by men.

Identifying gender via textual analysis. Personal letters closely approximate the conditions of face-to-face speech. Both linguistic transactions take place between two interlocutors who not only know one another but typically share personal experiences. Moreover, at least paradigmatically, both personal letters and conversation tend to be more informal than formal. We therefore expect (and find) a significant amount of “involved”

behavior in both instances (Biber, 1995), although female levels tends to be higher than those of males.

Does gender distinction carry over into other written venues? Mulac and Lundell (1994) studied impromptu essays that college students were asked to write describing landscape scenes that were projected onto a large screen. Drawing upon earlier work in the language and gender literature, researchers coded the essays with respect to 17 linguistic features, including “male language variables” (e.g., references to quantity, judgmental adjectives, elliptical sentences, locatives, and sentence-initial conjunctions or filler words) and “female language variables” (e.g., intensive adverbs, references to emotion, dependent clauses, sentence-initial adverbials, uncertainty verbs, hedges, and long mean length sentences). The authors found that the analysis of the essays with respect to gender-coded language variables correctly identified the essay-writer’s gender 72.5% of the time.

More recently, Argamon, Koppel, Fine, and Shimoni (2003) created an algorithm for identifying a writer’s gender by using computers to analyze written texts. The investigators compiled lists of linguistic features they judged to be indicative of female writers (e.g., use of pronouns) or of male writers (e.g., use of determiners, quantifiers, and cardinal numbers). The authors note that these feature sets align with Biber’s distinction between “involved” versus “informational” styles. Applying their algorithm to new samples, the algorithm correctly identifies the writer’s gender approximately 80% of the time (Koppel, Argamon, & Shimoni, 2002).

Standardized tests of written language skills. A very different venue for assessing gender differences in written language is standardized achievement tests. In the United States, the best-known yardstick of pre-college academic achievement is *The Nation’s*

Report Card, compiled by the National Assessment of Educational Progress (NAEP), under the auspices of the National Center for Educational Statistics. Over the years, girls have consistently outpaced boys on the writing component of the testing rubric. The 2002 study reports that for students tested in grades 4, 8, and 12, females continued to outscore their male counterparts, with the score gap between females and males being greatest in 12th grade. Evaluation was done using a six-point scale (ranging from “unsatisfactory” to “excellent”) that looked at each writing sample as a whole, rather than giving separate scores on individual dimensions (e.g., level of vocabulary, development of an argument, length, or writing mechanics).

Summary of writing and gender findings. We can look at our findings regarding writing and gender in terms of the two dichotomies we identified for analyzing speech and gender, namely social versus informative and standard versus non-standard. All of the corpus-based studies and the essay study support the speech-based finding that in general, female language is more social (“involved”) in character than is male. None of the studies we reviewed directly addressed the issue of standard versus non-standard language usage. However, the NAEP reports can be seen as providing supporting evidence for the general claim that the written language of females shows greater command of the “standard” than does the written language of males.

Gender and CMC

The field of computer mediated communication has sparked interest from sociolinguists concerned with the extent to which traditional gender dichotomies in face-to-face communication would play out in online behavior. Initial optimism that the relative anonymity of the medium would support an equalization of gender roles (e.g., Danet, 1998)

soon gave way to the realization that online dynamics often replicated off-line gender distinctions.

Herring (2003) offers a thorough analysis of language and gender issues in one-to-many CMC forums such as listservs and newsgroups (both of which involve asynchronous communication) and Chat, MUDs, and MOOs (all of which involve synchronous communication). In both venues, Herring reports gender asymmetries. On asynchronous discussion lists and newsgroups, “males are more likely to post longer messages, begin and close discussions in mixed-sex groups, assert opinions strongly as ‘facts,’ use crude language (including insults and profanity), and in general manifest an adversarial orientation toward their interlocutors” while females “tend to post relatively short messages, and are more likely to qualify and justify their assertions, apologize, express support of others, and in general, manifest an ‘aligned’ orientation toward their interlocutors” (Herring, p. 207).

In one-to-many synchronous CMC forums, gender roles are in some ways more balanced (e.g., more equal participation in Chat environments as measured by number of messages and message length; see Herring, 1999). However, Herring goes on to note that gender differences (and often inequalities) still pervade Chat and social MUDS or MOOs. For example, males use more aggressive and insulting speech acts, while females type three times as many representations of smiles or laughter. Male discourse is oppositional and adversarial, while female discourse style is aligned and supportive (Herring, 2003).

In considering the literature on one-to-many CMC gender issues, we need to be mindful that the social conditions for one-to-one CMC are quite different. Most of the one-to-many forums that have been studied are open to large numbers of users, the majority of

whom a given interlocutor does not personally know. Moreover, in many such forums, users can anonymize their postings, representing themselves as being of an age, gender, etc. that does not correspond to who they are in real life. In one-to-one venues such as IM, email, or texting on mobile phones, conversation typically takes place between interlocutors who know one another. IM conversations generally involve dyads who have ongoing face-to-face relationships that are supplemented by online discourse. Thus, while some of the gender findings from the body of one-to-many CMC literature may prove relevant for a gender analysis of IM language (e.g., use of emoticons representing laughing or smiling), other variables (e.g., dominance of a conversation) may be less applicable.

Given this background on the nature of IM as a form of CMC and on previous studies of language and gender, we now turn to the IM corpus upon which our own analysis is based.

Empirical Study of IM: Corpus, Terminology, Variables

An IM corpus was collected in April 2003 from 22 college-aged students who were attending school or had graduated the semester before the study was undertaken. This pilot study was part of a larger, ongoing project comparing online and face-to-face communication among American college students. The discussion that follows is restricted to a large subset of linguistic features for which data analysis has been completed.

The IM Corpus

Using America Online's freely downloadable program known as AIM (AOL Instant Messenger), IM conversations were initiated by a cohort of current (or recently graduated) students at American University in Washington, DC. American University is an academically competitive, largely residential institution with approximately 5500

undergraduates. These conversational initiators (“student experimenters”) were asked to IM specified numbers of male and female peers who were on the initiator’s AIM Buddy List. (A Buddy List is a set of IM screen names that a user chooses to associate with his or her account. When one of these users logs on to AIM, he or she can see which people on the Buddy List are currently online.)

At the beginning of each conversation, initiators requested permission of their interlocutor to save the IM conversation that was to follow so that it could be forwarded to the author for research purposes. Formal consent forms were distributed electronically to all parties (student experimenters and their conversational partners) at the end of the IM conversation. Both members of each conversation were given the opportunity to edit out any words or turns they wished to delete (an option rarely taken), and user screen names were anonymized. Student experimenters then electronically forwarded the IM conversation files (and consent forms) to a site established for project analysis. Initial discussion in the IM conversations regarding the research project was eliminated from actual data analysis, thus precluding an analysis of conversational openings.

The corpus consisted of 23 distinct IM conversations. Nine conversations took place between females (FF) and nine between males (MM). An additional five conversations involved female/male dyads (FM). In a number of the FF and FM conversations, a single student experimenter had conversations with several people on his or her Buddy List. Because several student experimenters withdrew from the project and could not be replaced, most of the MM conversations were between the same two interlocutors.

Taken collectively, the 23 IM conversations contained a total of 2185 conversational turns, made up of 11,718 words. Some of the analyses discussed below were performed on the entire corpus while others were restricted to comparison of the nine FF and nine MM conversations (together totaling 1861 conversational turns).

Terminology

As yet, there is no generally accepted linguistic terminology for analyzing IM data. Therefore, we established the following definitions.

Turn: composition (i.e., by typing) and transmission of an instant message

Max: hey man

Utterance: rough equivalent of a sentence in IM

Susan: Somebody shoot me!

Sequence: a number of IM turns in a row (from 1 to N) sent seriatim by the same interlocutor

Max: hey man

Max: whassup [sequence = 2 turns]

Utterance Chunking: breaking a single IM utterance into two or more turns

Joan: that must be nice

to be in love

in the spring

Closing: series of turns (between interlocutors) at the end of an IM conversation, beginning with one party initiating a closure of the conversation and ending with termination of the IM connection

Sam: Hey, I gotta go [first indication that will terminate conversation]

[subsequent conversational turns]

Sam: I'm outta here [final turn in conversation]

Linguistic Variables Analyzed

A range of linguistic variables were analyzed. The first three categories involved conversational scaffolding (essentially, how conversations were constructed). The first category was *turns*, analyzed in terms of average turn length (in words), longest turns (in words), percent of one-word turns (out of total turns), and turns per minute. Second, *sequences* were analyzed as longest sequence per conversation, average number of turns per sequence, and percent of sequences including >1 turn. Third, *conversation length* was analyzed for entire conversations (turns per conversation; time), and closings (number of turns; time).

We also considered a number of *lexical issues* that constitute shortenings of words or phrases (abbreviations, acronyms, and contractions) as well as looking at emoticons (i.e., graphic images that are intended to serve as sentence modifiers, expressing the sender's attitude towards the text preceding it). Tabulations of all linguistic variables were performed both for the entire corpus and by gender.

General Linguistic Profile of IM

The entire corpus was coded for each of the relevant variables, with instances of each variable then tallied and analyzed by hand. Despite some design and implementation limitations (including the overall size of the corpus and the skewed male sampling), the linguistic profile of IM that emerged clearly suggests a number of distinctive patterns in American college student IM conversations, as well as suggestive differences between male and female use of the medium. In this section, we will review the findings for the

overall corpus with respect to conversational scaffolding and lexical issues. In the following section, we consider these same domains with regard to gender.

Conversational Scaffolding

Turns. Table 1 summarizes the structural nature of individual turns.

INSERT TABLE 1 ABOUT HERE

The aggregated data suggest that average IM turns are fairly short (5.4 words per turn). However, averages obscure some of the important characteristics of contemporary college student IM conversations. One of these characteristics is that a number of IM turns are quite long indeed (the longest in the corpus being 44 words). A second characteristic is the high proportion of one-word turns (21.8% of the FF and MM combined corpus).

Since all of the turns in the IM conversations were time-stamped (a feature available directly through AIM), it was possible to calculate not only how long each conversation lasted but also how many turns took place per minute. The average—barely 4 turns per minute—is surprisingly low, considering how few seconds it takes to type an average of 5.4 words and transmit the message.

Sequences. For counting purposes, a sequence was defined as one or more consecutive terms from the same interlocutor. If a sender's multiple turn sequence was interrupted by a message from his or her interlocutor but the sender did not attend to the interruption, all consecutive turns dealing with the same theme were considered to be part of a single sequence. Table 2 summarizes the findings regarding sequences.

INSERT TABLE 2 ABOUT HERE

The first step in the analysis was to cluster all the data into sequences. There were 1292 sequences, built out of a total of 2185 turns. (Recall that “sequence” is here defined as one or more consecutive turns, so single turns are included in this statistic.) Sequences ranged in length from 1 to 18 turns. The average number of turns per sequence was 1.7. However, 42% of the IM sequences contained more than one turn (543 multi-turn sequences out of 1292 total sequences). If we eliminate the one-word turns (which constitute 21.8% of the total conversational turns but are rarely part of multi-turn sequences), multi-turn sequences constitute nearly half of the IM discourse.

Conversation length. Table 3 presents findings on number of turns and chronological length of both entire conversations and closing sequences.

INSERT TABLE 3 ABOUT HERE

Aggregated data suggest that IM conversations are fairly lengthy: over 93 turns apiece on average and nearly 24 minutes long. In reality, IM conversations show enormous variety, ranging from quick three or four turn volleys to conversations stretching over more than 200 turns and well over an hour. Moreover, although the communication channel may technically be open for an extended period of time, the two interlocutors are not necessarily directly engaging with one another over the entire IM conversational span. For example, in one lengthy FF conversation (142 turns, 88 minutes), there was a 15 minute gap when no turns occurred.

The data on IM conversational closings suggest that much as in real-life encounters, interlocutors who know one another reasonably well often take awhile before actually

terminating a conversation. Here is an example of a typical multi-turn IM conversational closing between female interlocutors. The sequence took 19 seconds:

Gale: hey I gotta run
 Sally: Okay.
 Sally: I'll ttyl?
 Gale: gotta do errands.
 Gale: yep!!
 Sally: Okay.
 Sally: ☺
 Gale: talk to you soon
 Sally: Alrighty.

The median closing sequence took 7 turns and averaged nearly 32 seconds.

Lexical Issues

We turn now to the four lexical issues examined in this study: abbreviations, acronyms, contractions, and emoticons. As we will see, the data argue for a somewhat different profile of IM (at least as used by American college students) than the image presented in the popular press of a medium filled with lexical short-cuts and emoticons, not to mention chaotic punctuation and devil-may-care spelling.

Abbreviations. Table 4 summarizes the use of what we might call CMC abbreviations (i.e., abbreviations that appear to be distinctive to CMC communication). Excluded from this tabulation are abbreviations that, while appearing in CMC messages, are also part of common off-line written usage (e.g., *hrs = hours*) or are direct representations of spoken usage (e.g., *cuz = because*). Admittedly, the line between

common off-line and CMC-specific usage is sometimes difficult to draw (e.g., *b/c* for *because* was included in the tally of CMC abbreviations while *prob* for *problem* or *convo* for *conversation* was not).

INSERT TABLE 4 ABOUT HERE

Considering that the IM corpus contained nearly 12,000 words (i.e., tokens), the fact that only 31 CMC abbreviations appeared (less than .3% of the corpus) is noteworthy.

Acronyms. Table 5 summarizes the use of what we might call CMC acronyms, i.e., acronyms that appear to be distinctive to CMC communication. Excluded from this tabulation are acronyms that, while appearing in CMC messages, are also part of common off-line written usage (e.g., *US* = *United States* or *TA* = *teaching assistant*).

INSERT TABLE 5 ABOUT HERE

As in the case of CMC abbreviations, the limited number (and variety) of CMC acronyms is small (90 tokens, which represents less than .8% of the total words in the corpus). The most frequent acronym used was *lol* (*laughing out loud*), an acronym that has even found its way into spoken usage among some college students. Interestingly, however, the IM (or even spoken) term is not always used to indicate the humorous response suggested by the words “laughing out loud”. Rather, both *lol* and *heehee* (or *haha*) are commonly used as phatic fillers for the equivalent of *OK*, *cool*, or *yeah*. Thus, we find a contrast between “literal” CMC meaning

Laura: What’s your favorite dish to make?

Cynthia: hehehe
 Cynthia: chocolate mousse
 Laura: freedom fries?
 Cynthia: LOL

and phatic meaning

Mark: i've got this thing that logs all convos [=conversations] now
 Jim: really?
 Jim: why's that
 Mark: i have ever [=every] conversation i've had with anybody since the
 16th
 Mark: i got a mod [=module] for aim [=AIM], and it just does it
 Mark: i'm not sure why
 Jim: lol
 Jim: cool

Contractions. Much as popular conceptions of IM lead us to anticipate a high frequency of textual shortening through use of abbreviations and acronyms, we might expect writers to make use of contractions wherever the language permits (e.g., *I'm* instead of *I am* or *he's* rather than *he is*). However, just as the college IM corpus yielded surprisingly few CMC abbreviations or acronyms, an analysis of tokens representing potential contractions suggests that again, many users are not taking advantage of possible lexical shortcuts. Table 6 summarizes the occurrences of contracted versus uncontracted lexical pairs occurring in the corpus:

 INSERT TABLE 6 ABOUT HERE

Contracted forms constitute the majority of usages of contraction pair choices. In 65.3% of situations, the contracted option was chosen. However, the finding that in nearly 35% of the linguistic situations offering an option, users chose the uncontracted “long form” (e.g., *I am, he is, do not, they are*) was unexpected.

Emoticons. Table 7 summarizes all types and tokens of emoticons appearing in the full corpus of nearly 12,000 words.

 INSERT TABLE 7 ABOUT HERE

Given the amount of publicity that emoticons continue to receive in the popular press (e.g., Lorenzi, 2002), it was surprising to see how few emoticons appeared in the entire text. Of the 49 emoticons used, 31 of them (nearly two-thirds) were a “smiley”. Moreover, a small number of subjects were responsible for using the majority of 49 emoticons. Out of 22 interlocutors, just three people accounted for 33 of the emoticons in the corpus. This pattern is consistent with Walther and D’Addario’s (2001) finding in an email experiment that the emotional meaning emoticons convey is overshadowed by the verbiage accompanying them.

Gender Issues in IM

The IM corpus was next examined with respect to gender. For conversational scaffolding, comparisons were made for turn structure and conversation length. (Sequence data were not tabulated data by gender.) For lexical issues, gender comparisons were done for abbreviations, acronyms, contractions, and emoticons. No gender-based patterns

emerged for abbreviations or acronyms, and therefore neither issue will be discussed in this section. However, in the case of both contractions and emoticons, gender appears to be a relevant variable.

Conversational Scaffolding

Turns. Table 8 summarizes the gender comparison of the 9 female-female (FF) versus the 9 male-male (MM) IM conversations with regard to turn scaffolding.

INSERT TABLE 8 ABOUT HERE

On most measures involving turns (average turn length in words, percent of one-word turns, and average turns per minute), there appears to be little difference between female and male usage patterns. The only domain in which genders showed obvious variation is longest turns. For the FF and MM instant messaging conversations analyzed, if we consider just the longest turn in each conversation (i.e., the longest turn in each of the 9 FF conversations; the longest turn in each of the 9 MM conversations), we find that female “longest turns” (27.6 words) were an average of 8 words lengthier than male “longest turns” (19.6 words). Moreover, the longest single turn in all of the FF conversations was 44 words, while the longest single turn in all of the MM conversations was 34 words.

Conversation length. Table 9 presents the gender breakdown regarding conversation length issues.

INSERT TABLE 9 ABOUT HERE

Gender seems to make a difference with regard both to overall conversation length (in number of turns and clock time) and to the length of conversational closings. With regard to overall conversation length, FF conversations were roughly a third longer (in turns and time) than were MM conversations, although the considerable range in turns per conversation makes direct gender comparison difficult. When it came to closing IM conversations, however, females clearly used a greater number of turns, and lengthier time spans, than did males.³

Lexical Issues

Contractions. Table 10 summarizes the use of contractions in the FF and MM corpus, by gender.

INSERT TABLE 10 ABOUT HERE

Both males and females used higher percentages of contracted than uncontracted forms in their IM conversations (e.g., using *don't* rather than *do not*). However, males used a much greater proportion of contracted forms than did females.⁴ While males chose contracted forms more than three-quarters of the time (77.1%), females were more evenly balanced in their choices, using contracted forms only 57% of the time.

Emoticons. Table 11 summarizes the use of emoticons by, gender, for the entire IM corpus.

INSERT TABLE 11 ABOUT HERE

A greater proportion of females used emoticons than did males.⁵ Of the 16 female participants in the study, three-quarters used one or more emoticons in their IM conversations. Of the 6 male participants, only one used emoticons. As for total emoticon usage by gender, 34 (of the total 49) emotions were used by females and 15 by males. However, all of the 15 male usages are attributable to a single male interlocutor, and all 15 occurred when that male was in IM conversations with female interlocutors.

IM and Gender: Discussion and Future Research

General Linguistic Profile and Interpretation

Our analysis of a corpus of instant messaging conversations constructed by American college students offers an initial linguistic profile of IM practices by a subset of contemporary users. For the database as a whole, we found that with respect to conversational scaffolding, turns vary in length and frequency, but average 5.4 words per turn and 4 turns per minute. The use of multi-turn sequences by interlocutors is common, constituting 42% of the corpus. Conversations vary widely in length and typically take multiple turns to close. Looking at the entire corpus with respect to lexical issues, we saw that the use of CMC abbreviations and CMC acronyms was minimal, with *lol* by far the most common acronym. In lexical situations where either a contracted or uncontracted form could be used, interlocutors used contracted forms only two-thirds of the time. The use of emoticons was minimal, with the smiley by far the most common.

The overall linguistic analysis provides evidence that IM conversations, at least as constructed by the present American college student sample, represent a blend of both spoken and written language conventions. As for speech, the conversations make relatively frequent use of the acronym *lol* as a phatic filler, roughly comparable to *OK*, *really*, or *yeah*

in spoken discourse. Similarly, interlocutors took multiple turns to close IM conversations, a feature familiar in face-to-face discourse.

As for writing, interlocutors did not regularly avail themselves of the common “shorthand” techniques popularly associated with IM. Not only were CMC abbreviations and acronyms infrequent, but even contractions were not as pervasive as might be expected in a medium touted as favoring brevity. As was reported elsewhere (Baron & Ling, 2003), this IM corpus revealed other signs of a more formal written style, including a paucity of spelling errors and presence of a number of attempts to correct one’s own typing errors made in an immediately preceding turn.

Equally striking was the preponderance of multi-turn sequences. While motivations for breaking single utterances into multiple transmissions still need to be formally investigated, anecdotal evidence from students suggests three plausible explanations. The first is technological. By transmitting pieces of a message seriatim, senders provide some text for their interlocutors to begin reading (while the sender is typing more of the message). This piecemeal transition style keeps the interlocutors from waiting and, equally importantly, maintains the interlocutors’ attention so they don’t shift to another task (e.g., participating in an IM conversation with someone else, surfing the Web, making a phone call, etc.). The second motivation may have to do with layout of text on the page. It is easier (and faster) to read short lines of text than to read long lines spanning the entire page horizontally. The final motivation may be aesthetic. Several IM users have reported that when they divide IM utterances into chunks across multiple turns, they are consciously attempting to make the results visually resemble a poem.

Perhaps the most unexpected finding was how few average turns (4) were exchanged per minute. With turns averaging 5.4 words, that means an average of 21.6 words were sent per minute, which probably represents considerably less than half the typing speed of most college students. Anecdotal evidence indicates that the majority of college students engage in multitasking while doing IM. Given the time gaps between many turns, we can only conclude that in many instances, this “synchronous” form of CMC is being used asynchronously.

Gender Findings and Interpretation

When we reexamined the corpus looking for gender-based patterns of linguistic usage, we found several aspects of conversational scaffolding in which male and female patterns differed. Whereas average turn length was similar across genders, the longest single turns in FF conversations were longer than those in MM conversations. The average FF conversation was longer (both in number of turns and in time) than the average MM conversation. Females took roughly twice as long (in number of turns and in time) to close conversations as did males. With regard to lexical issues, females used more uncontracted lexical forms than did males, and females were far more likely to use emoticons than were males.

We can think about these gender findings in terms of the two broad categories we earlier identified for looking at language and gender issues, namely social/involved versus informative, and standard versus non-standard.

Social/involved versus informative. All of the IM conversational scaffolding findings indicate that female interlocutors were more “talkative” than their male counterparts. Females took longer terms, had longer overall conversations, and took longer

to say goodbye. In our earlier review of the sociolinguistic literature, we found no grounds for arguing that even in same-sex dyadic face-to-face conversation, women are more talkative than men. However, in their study of informal essay-writing, Mulac and Lundell (1994) reported that females used longer sentences than did males. Thus, it is possible that our IM findings reflect a female *writing* style rather than a female *speech* style.

It is instructive to compare the IM gender findings with those reported for one-to-many CMC communication. In the lexical domain, our finding that females were far more likely to use emoticons than were males is consonant with Herring's report (2003) that women were three times as likely to use representations of smiles or laughter than were males in one-to-many synchronous communication. However, our IM findings with regard to verbosity were not paralleled in one-to-many CMC, for which Herring (1999, 2003) reported men used longer messages in asynchronous communication and more balanced numbers and length of messages in synchronous CMC. Herring's findings more closely parallel those summarized by Holmes (1993) for public spoken conversations among mixed-sex participants.

Standard versus non-standard. Our finding that females used fewer contracted forms than did males suggests that females have a greater tendency towards treating IM as a written medium (since in the academic domain, students are still taught that contractions should not be used in formal writing). The number of CMC abbreviations and acronyms in the IM corpus was too small for drawing any gender-based distinctions. Preliminary (though incomplete) analysis of punctuation and capitalization in the IM conversations suggests that females are more normative than are males in that females employed more standard punctuation and capitalization than did their male counterparts. Other studies of

one-to-one CMC such as Ling's work (in press) on Norwegian short text messaging report that female teenagers and young adults used more standard punctuation and capitalization in their text messaging on mobile phones than did males.

Future Research

The results reported in this paper represent a first step towards understanding the issue of gender in instant messaging conversations. However, additional data collection and analysis remain to be done before we can be more definitive about the linguistic nature of contemporary IM. First, IM data need to be collected from a broader range of cohorts (including college students from other types of academic settings, teenagers, adults in the work force, etc.). Second, IM data need to be coded for a variety of additional variables. With respect to the social versus informative dimension of language use, we need to look at such issues as content analysis (cf. Boneva & Kraut, 2002 on content analysis of email), pronoun use (cf. reports by Argamon et al., 2003; Biber et al., 1998 that females use more pronouns than males), and use of apologies (cf. Herring, 2003 regarding women's heavier use of apologies in one-to-many CMC). With respect to the standards issue, we need to look closely at punctuation and capitalization, as well as at spelling errors and error correction. That is, we need to study the IM data more extensively to determine the extent to which gender distinctions previously reported in spoken and written language are reflected in IM conversations.

Third, we need to study the nature of multitasking behavior while engaging in IM conversations. We need to look for possible gender distinctions here, since anecdotal evidence indicates that males may be heavier multitaskers than females. Data on multitasking will provide a better understanding of the extent to which IM should be

thought of as a form of synchronous conversation as opposed to being a potentially less engaged form of communication. And fourth, to better understand the issue of comparative female verbosity in IM conversations, it will be important to collect both dyadic speech data and off-line written compositions from comparable cohorts so that we can judge with better certainty the extent to which male and female IM conversations approximate spoken or written discourse. Existing gender studies of speech or writing may not reflect the language patterns of contemporary college students, and therefore may not make for appropriate comparisons with college student IM conversations.

Conclusions

Although the IM conversations of American college students reveal considerable variation across users, as a whole, they evidence a low proportion of the characteristics assumed in the popular press. Analysis of average message length and average turns per minute suggests that IM is frequently used asynchronously, in tandem with multitasking behaviors, though the precise character of such multitasking remains to be studied.

Unlike one-to-many forms of CMC, IM conversations contain gender patterns that in some ways reflect gender distinctions in spoken and (off-line) written language. Our findings that females use longer IM turns and fewer contractions than males suggest that female IM conversations reflect more a female writing style than a female speech style, though comparative samples of face-to-face conversation and of writing by college-aged students will need to be collected and analyzed before this hypothesis can be confirmed.

REFERENCES

- Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text, 23*, 321-346.
- Baron, N. S. (1984). Computer mediated communication as a force in language change. *Visible Language, 18*, 118-141.
- Baron, N. S. (1998). Letters by phone or speech by other means: The linguistics of email. *Language and Communication, 18*, 133-170.
- Baron, N. S. (2000). *Alphabet to email: How written English evolved and where it's heading*. London and New York: Routledge.
- Baron, N. S. (2003). Language and the internet. In A. Farghali (Ed.), *The Stanford handbook for language engineers* (pp. 59-127). Stanford, CA: CSLI Publications.
- Baron, N. S. & Ling, R. (2003, October). *IM and SMS: A linguistic comparison*. Paper presented at the Fourth International Conference of the Association of Internet Researchers, Toronto, Ontario, Canada.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of Register Variation*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D. & Finegan, E. (1997). Diachronic relations among speech-based and written registers in English. In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the*

- present: Studies in the changing English language in honour of Matti Rissanen* (pp. 253-275). Helsinki: Modern Language Society.
- Boneva, B., & Kraut, R. (2002). Email, gender, and personal relations. In B. Wellman & C. Haythornthwait (Eds.), *The Internet in everyday life* (pp. 372-403). Oxford: Blackwell.
- Boneva, B., Quinn, A., Kraut, R., Kiesler, S., & Shklovski, I. (in press). Teenage communication in the instant messaging era. In R. Kraut, M. Brynin, & S. Kiesler (Eds.), *Information technology at home*. Oxford: Oxford University Press.
- Cameron, D. (1998). Gender, language, and discourse: A review essay. *Signs: Journal of Women in Culture and Society*, 23, 995-973.
- Chambers, J. K. (1992). Linguistic correlates of gender and sex. *English World-Wide*, 13, 173-218.
- Collott, M., & Belmore, N. (1996). Electronic language: A new variety of English. In S. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 13-28). Amsterdam: John Benjamins.
- Coates, J. (1993). *Women, men, and language* (2nd ed.). London and New York: Longman.
- Coates, J. (Ed.) (1998). *Language and gender: A reader*. Oxford: Blackwell.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Danet, B. (1998). Text as mask: Gender and identity on the Internet. In S. Jones (Ed.), *Cybersociety 2.0* (pp. 129-158). Thousand Oaks, CA: Sage.
- Eckert, P. (1989). *Jocks and burnouts: Social categories and identities in high school*. New York: Teachers College Press.

- Eckert, P., & McConnell-Ginet, S. (2003). *Language and gender*. New York: Cambridge University Press.
- Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8, 8-34.
- Grinter, R. E., & Palen, L. (2002). Instant messaging in teen life. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '02)*, New Orleans, LA, November 16-20. New York: ACM Press.
- Hård af Segerstad, Y. (2002). *Use and adaptation of written language to the conditions of computer-mediated communication*. Doctoral Dissertation, Department of Linguistics, Göteborg University.
- Helderman, R. (2003, May 20). Click by click, teens polish writing. *Washington Post*, B1.
- Herring, S. (Ed.) (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. Amsterdam: John Benjamins.
- Herring, S. (1999). The rhetorical dynamics of gender harassment online. *The Information Society*, 15, 151-167.
- Herring, S. (2002). Computer-mediated communication and the Internet. In B. Cronin (Ed.), *Annual review of information science and technology* 36 (pp. 109-168). Medford, NJ: American Society for Information Science and Technology.
- Herring, S. (2003). Gender and power in on-line communication. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 202-228). Malden, MA: Blackwell.
- Holmes, J. (1993). Women's talk: The question of sociolinguistic universals. *Australian Journal of Communication*, 20, 125-149.

- Holmes, J. (1995). *Women, men, and politeness*. New York: Longman.
- Holmes, J., & Meyerhoff, M. (Eds.) (2003). *The handbook of language and gender*. Malden, MA: Blackwell.
- Jacobs, G. (2003, April). *Breaking down virtual walls: Understanding the real space/cyberspace connections of language and literacy in adolescents' use of instant messaging*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- James, D. (1996). Women, men, and prestige speech forms: A critical review. In V. L. Bergvall, J. M. Bing, & A. F. Freed (Eds.), *Rethinking language and gender research* (pp. 98-125). London: Longman.
- Jones, S. (2002). The Internet goes to college: How students are living in the future with today's technology. *Pew Internet & American Life Project*. Retrieved July 25, 2004, from <http://www.pewinternet.org/reports/toc.asp?Report=71>.
- Kendon, A. (1980). The sign language of the women of Yuendumu: A preliminary report on the structure of Warlpiri sign language. *Sign Language Studies*, 27, 101-112.
- Koppel, M., Argamon, S., & Shimoni, A. R. (2002). Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17, 401-412.
- Labov, W. (1991). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205-254.
- Lee, J. (2002, Sept. 19). I think, therefore IM. *New York Times*, Circuits Section, p. G1.
- Lenhart, A., Rainie, L., & Lewis, O. (2001). Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family

- relationships. *Pew Internet & American Life Project*. Retrieved July 25, 2004, from http://www.pewinternet.org/pdfs/PIP_Teens_Report.pdf.
- Ling, R. (in press). The sociolinguistics of SMS: An analysis of SMS use by a random sample of Norwegians. In R. Ling & P. Pedersen (Eds.), *Front-stage/back-stage: Mobile communication and the renegotiation of the social*. London: Springer-Verlag.
- Lorenzi, R. (2002). 'Smiley' emoticon turns 20. Retrieved July 25, 2004, from <http://dsc.discovery.com/news/briefs/20020916/emoticon.html>.
- Maynor, N. (1994). The language of electronic mail: Written speech? In G. D. Little & M. Montgomery (Eds.) *Centennial usage studies* (pp. 48-54). Tuscaloosa, AL: University of Alabama Press.
- Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language and Communication* 14, 299-309.
- Nardi, B., Whittaker, S., & Bradner, E. (2000). Interaction and outeraction: Instant messaging in action. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '00)*, Philadelphia, PA (pp. 79-88). New York: ACM Press.
- National Center for Educational Statistics (2002). *The nation's report card: Writing 2002*. National Assessment of Educational Progress. Washington, DC: U.S. Department of Education, Institute of Educational Sciences.
- Palander-Collin, M. (1999). Male and female styles in seventeenth century correspondence. *Language Variation and Change*, 11, 123-141.

- Randall, N. (2002). *Lingo online: A report on the language of the keyboard generation*. Department of English, University of Waterloo. Retrieved July 25, 2004, from <http://www.arts.uwaterloo.ca/~nrandall/LingoOnline-finalreport.pdf>.
- Romaine, S. (2003). Variation in language and gender. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (pp. 98-118). Malden, MA: Blackwell.
- Sachs, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- Schegloff, E., & Sachs, H. (1973). Opening up closings. *Semiotica*, 7, 289-327.
- Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds., 2001). *The handbook of discourse analysis*. Malden, MA: Blackwell.
- Schiano, D., Chen, C., Ginsberg, J., Gretarsdottir, U., Huddleston, M., & Isaacs, E. (2002). Teen use of messaging media. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '02)*, Minneapolis, MN (pp. 594-595). New York: ACM Press.
- Tannen, D. (1994). *Gender and discourse*. New York: Oxford University Press.
- Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language? *Psychological Science*, 12, 171-175.
- Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review*, 19, 323-345.
- Yates, S. (1996). Oral and written linguistic aspects of computer conferencing. In S. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-*

cultural perspectives (pp. 29-46). Amsterdam: John Benjamins.

NOTES

1. The author is grateful to Joe Walter for his valuable input to the final version of this article.

2. MUDs and MOOs are acronyms for Multiple User Dungeons (or Dimensions or Discussions – the term has evolved over the years) and MUDs Object Oriented, respectively. The term “Object-Oriented” refers to a programming style.

3. Large variances associated with the means on turns per conversations (FF $M = 121.89$, $SD = 82.68$; MM $M = 85.22$, $SD = 88.96$) and on minutes per conversation (FF $M = 31.33$, $SD = 25.57$; MM $M = 18.89$, $SD = 17.27$) rendered no statistically significant differences between gender pairs. For IM closings, however, FF/MM differences were significant both for number of turns per closing (FF $M = 9.78$, $SD = 5.12$; MM $M = 4.29$, $SD = 1.13$), $t(14) = 2.77$, $p = .015$; as well as seconds taken to close (FF $M = 41.00$, $SD = 20.12$; MM $M = 16.29$, $SD = 15.48$), $t(14) = 2.68$, $p = .018$.

4. Fischer’s exact test $p = .022$.

5. $\chi^2(1) = 50.44$, $p < .001$.

Table 1.

Turns Summary

Average turn length (in words):	5.4
Longest turn (in words):	44.0
Percent of one-word turns (out of total turns):	21.8%
Average turns per minute:	4.0

Note. Average turn length and turns per minute were calculated over the entire data base of 2185 turns. However, longest turns and percent of one-word turns were calculated over just the FF and MM conversations (totaling 1864 turns).

Table 2.

Sequences Summary

Number of sequences in total data base:	1292
Longest IM sequence per conversation (i.e., from same interlocutor)	18 sequential turns
Average number of turns per IM sequence:	1.7
Percent of IM sequences including >1 turn:	42.0%

Table 3.

Conversation and Closing Length

<i>Entire Conversations</i>	<i>Mean</i>
Average turns per conversation:	93.8 turns
Average time per conversation:	23.7 minutes

<i>Closings</i>	<i>Mean</i>
Average number of turns to close:	7.0 turns
Average time to close:	31.9 seconds

Table 4.

CMC Abbreviations

Total CMC abbreviations:	31 (out of 11,718 words)
Type/token distribution (arranged alphabetically):	
<i>bc</i> (also <i>b/c</i>) = <i>because</i>	5
<i>bf</i> = <i>boyfriend</i>	2
<i>cya</i> = <i>see you</i>	7
<i>k</i> = <i>OK</i>	16
<i>y?</i> = <i>why</i>	1

Table 5.

CMC Acronyms

Total acronyms:	90 (out of 11,718 words)
Type/token distribution (arranged alphabetically)	
<i>brb = be right back</i>	3
<i>btw = by the way</i>	2
<i>g/g (also g2g) = got to go</i>	2
<i>LMAO = laughing my <xxx> off</i>	1
<i>lol (also LOL) = laughing out loud</i>	76
<i>OMG = Oh my God</i>	1
<i>ttyl = talk to you later</i>	5

Table 6.

Use of Contracted versus Uncontracted Lexical Pairs

Contracted:	498
Uncontracted:	265
Total:	763
% Contracted (of total):	65.3%

Note. This analysis is based upon the 9 FF and 9 MM conversations.

Table 7.

Emoticon Summary

Total emoticons: 49 (out of 11,718 words)

Type/token distribution (by descending frequency)

:-) = smiley	31
:-(= frowny	5
O:-) = angel	4
:-P = sticking out tongue, with nose	3
;-) = winking	2
:-\ = undecided	1
:-[= embarrassed	1
:P = sticking out tongue, without nose	1
:- = [probably a typographical error]	1

Table 8.

Turns Summary by Gender Pairs

	<i>Gender Pairs</i>	
	FF	MM
Average turn length (in words)	5.3	5.2
Longest turns (in words)		
averaged over 9 conversations	27.6	19.6
absolute longest	44	34
One-word turns as percent of total	20.0%	24.4%
Average turns per minute	3.9	4.5

Note. While Table 8 tabulates only the nine FF and the nine MM conversations, the earlier turns summary in Table 1 calculates average turn length and turns per minute over the entire data base, but calculates longest turns and percent of one-word turns just over the FF and MM conversations.

Table 9.

Conversation and Closing Length by Gender Pairs

	<i>Gender Pairs</i>	
	FF	MM
<i>Entire Conversations</i>		
Average turns per conversation:	121.9	85.2
Average minutes per conversation:	31.3	18.9
<i>Closings</i>		
Average number of turns to close:	9.8	4.3
Average seconds to close:	41.0	16.3

Table 10.

Use of Contracted versus Uncontracted Lexical Pairs by Gender

	<i>Gender</i>	
	FF	MM
Contracted:	256	242
Uncontracted:	193	72
Total:	449	314
% Contracted (of total)	57.0%	77.1%

Table 11.

Emoticon Usage by Gender

	Females	Males
	(<i>n</i> = 16)	(<i>n</i> = 6)
Number of interlocutors using emoticons:	12	1
Total emoticons used:	34	15
