LINGUISTIC RUIN? LOL!
INSTANT MESSAGING
AND TEEN LANGUAGE

SALI A. TAGLIAMONTE
DEREK DENIS
University of Toronto

ABSTRACT: This article presents an analysis of Instant Messaging (IM), a one-to-one synchronous medium of computer-mediated communication. Innumerable articles in the popular press suggest that increasing use of IM by teens is leading to a break-down in the English language. The analyses presented here are based on a unique corpus involving 72 teenagers and over a million words of natural, unmonitored IM. In addition, a corpus of speech from the same teenagers is examined for comparison. Targeting well-known IM features and four areas of grammar, we show that IM is firmly rooted in the model of the extant language. It reflects the same structured heteroge-neity (variation) and the same dynamic, ongoing processes of linguistic change that are currently under way in contemporary varieties of English. At the same time, IM is a unique new hybrid register, exhibiting a fusion of the full range of variants from the speech community—formal, informal, and highly vernacular.

TEENAGERS IN THE EARLY TWENTY-FIRST CENTURY are using home com-puters for communication at unprecedented rates in ever-expanding virtual communities. A particularly favorite medium, at least when we conducted this research, was Instant Messaging (IM). IM is "a one-to-one synchronous form of computer-mediated communication" (Baron 2004, 13). It is "direct, immediate, casual online contact" (Schiano et al. 2002). In essence, IM is real-time "interactive written discourse" (Ferrara, Brunner, and Whittemore 1991, 8).

Examples of IM in (1) show snippets of conversations in the IM environ-ment. The numbers in square brackets indicate the individual (i.e., in 1a [025] converses with [3], and in 1b [2] converses with [006]).[1]

1. a. [025] lol y dont i believe u
        [3] its trueeee :P well fine then
        [3] what have YOU been doing
        [025] i dont kno nothing really
        [TEEN/025, male, 17; TEEN/3, female, 17][2]

    b. [2] when i drink tea
        [oo6] ew u do!?
        [oo6] hahahha
        [2] yes b/c it tastes good
        [2] with milk tea
        [2] so it's sugary
        [2] actually i drink tea with chocolate powder
        [2] it's the best thing ever
        [oo6] i'll try it next time...
        [2] YOU SHOULD TRY IT
        [TEEN/oo6, female, 17; TEEN/2, female, 17]

Innumerable articles in the popular press have targeted IM. They suggest that it is leading to a "breakdown in the English language," "the bastardization of language" (O'Connor 2005), even "the linguistic ruin of [the] generation" (Axtman 2002). In contrast, linguists argue that it is not the result of students' lax attitude toward spelling and grammar, but characteristic of a general "linguistic whatever-ism" (Baron 2003a, 5). Indeed, some have suggested that discourse on the Internet is a "new species of communication," complete with its own lexicon, graphology, grammar, and usage conditions (Crystal 2001, 48).

In this study, we add to the building information on computer-mediated communication (CMC) in general and the nature of IM language in particular by setting out to discover what IM language is like among the sector of the population that uses it the most—teenagers. We approach this task from two entirely different perspectives. First, we respond to the prevailing lay perception that IM language is riddled with abbreviations, short forms, and symbolic uses. The media presentation of these features suggests that they infuse IM language, but how prevalent are they? Second, because one of our goals is to contribute new information to the task of placing IM on the written-to-spoken linguistic spectrum, we conduct a distributional analysis of a feature identified as one of the "main differences between speech and writing" (Yates 1996, 40)—the way individuals reference themselves and others using personal pronouns (see, e.g., Chafe 1982). Finally, we offer a new perspective on IM language by conducting an analysis of structured heterogeneity in the linguistic system (Weinreich, Labov, and Herzog 1968, 99–100). Such systems may or may not operate the same way in a medium such as IM as they do in spoken language.[3] We draw on the findings of earlier studies to inform our selection of variable features, focusing on those that contrast different types of change in contemporary English because we hypothesize this will give us unique insight into the nature of IM. Rapidly innovating features among adolescents, such as intensifier *so* (Tagliamonte forthcoming) and quotative *be like* (Tagliamonte and D'Arcy 2005, 2007a),

will enable us to determine if IM language is on the forefront of linguistic change. Longitudinal changes, such as the rise of the *going to* future (e.g., Poplack and Tagliamonte 1999) and deontic *have to* (Tagliamonte and Smith 2006; Tagliamonte and D'Arcy 2007b), will enable us to assess the extent to which IM language mirrors everyday language. We also contribute a novel methodological approach to the study of CMC by utilizing comparative sociolinguistic methods. These techniques enable the analyst to model linguistic variation and change in related corpora (e.g., Tagliamonte 2002, 2006). Taken together, our analyses lead to an unexpected and intriguing finding. IM language is characterized by a robust mix of features from both informal spoken registers and more formal written registers—in essence it is a hybrid register.

In subsequent sections, we provide a brief history of CMC and IM and review the relevant literature, describe our data and methodology, present the linguistic analyses, and offer our conclusions.

COMPUTER-MEDIATED COMMUNICATION: HISTORY AND RESEARCH

Computer-mediated communication has developed at an astonishing rate in its short history.[4] With the launch of the World Wide Web in 1990 and the mass popularization of the Internet, many forms of CMC became widely used. These different types of CMC can be differentiated on two parameters: (1) the number of recipients of a message and (2) the synchronicity of the communicative event. Though it is possible to have multiple recipients in IM, it is primarily used for one-to-one dialogue. IM is also synchronous, since participants are "electronically present at the same time" (Paolillo 1999, 2). Instant (or near-instant) reply is the norm, and spontaneous, real-time dialogue takes place. IM is unique in that it is the only one-to-one synchronous type of CMC.

At the turn of the twenty-first century, 80% of Canadian teens and 74% of American teens had used an IM program (Lenhart, Rainie, and Lewis 2001, 3; Randall 2002, 21). IM's skyrocketing popularity among young people has raised concern from a variety of people, including parents, teachers, psychologists, and grammarians. This has led to an explosion of studies on the topic, with more arising every day.

Nearly all types of CMC have been subject to academic study, many of which appear in the online *Journal of Computer-Mediated Communication* (1995–, ed. Susan Herring, http://jcmc.indiana.edu). Others appear independently in journals from many fields, including linguistics, computer science, sociology, and psychology. Unfortunately, as Cherny (1999, 2) states, "research

on . . . CMC has occurred on the fringes of many disciplinary traditions, often making it difficult to find previous work on the topic." Since few studies have singled out IM in particular, relevant research on all forms of CMC will be discussed in order to lay the foundation of our study. The majority of research focuses on providing a descriptive analysis of CMC's "funky" (Palfreyman and al Khalil 2003) features (Werry 1996; Hentschel 1998; Randall 2002), although some studies conduct discourse and/or linguistic analysis (Ferrara, Brunner, and Whittemore 1991; Yates 1996; Cherny 1999; Thurlow 2002; Baron 2003a, 2004). Table 1 summarizes the empirical linguistic research on CMC. It comprises a considerable range of different research foci, including

TABLE 1
Summary of Linguistic Computer-Mediated Communication Research

Year	Researcher(s)	Form of CMC	Corpus Size	Feature
1991	Ferrara et al.	e-messages (an early form of IM)	18,769 words	subject drop article deletion copula deletion tersed sentences shortened words
1996	Yates	computer conferencing (early newsgroups)	Written: Lancaster-Oslo/Bergen; Spoken: London Lund; CMC: 2,222,049 words	type/token ratio analysis pronoun use modal auxiliaries
1996	Werry	IRC	no corpus	conversation organization addressivity abbreviation prosody
1998	Hentschel	IRC	no corpus	prosody lexicon orthography turn taking
1999	Paolillo	IRC	6,317 lines	use of r, u, z use of code-switching, Hindi-English use of obscenity
2003	Palfreyman & al Khalil	IM (MSN)	2,400 Arabic words & 2,000 English words	use of Roman keyboard set to write in Arabic use of local dialect
2004	Baron	IM (AIM)	11,718 words	turn taking conversation length openings & closings abbreviations contractions emoticons

features of prosody, spelling variants, and dialect representations. In the next section, we summarize the major findings of these studies, focusing on the main research question in this literature—what type of register is IM?

Ferrara, Brunner, and Whittemore (1991, 22–23) were among the first to address the question of whether CMC media are more like speech or more like writing. To collect their data, an experiment was created in which 23 subjects used e-messages (an early form of IM) to exchange messages with a controlled user who helped organize subjects' travel plans. Though messages contained features often associated with speech, rather than writing, well-known features of written registers occurred as well (24).

Yates's (1996, 46) groundbreaking study compared computer conferencing data from corpora of written and spoken language constructed by other researchers for other purposes. The results suggest a "neither simply speech-like nor simply written-like" assessment of CMC. While CMC was found to be similar to written discourse based on type/token ratio tests and lexical density, he found that the use of pronouns and modals was similar to speech (Yates 1996, 37, 39, 41, 45). Under the assumptiong that CMC has similar effects cross-linguistically, additional research supports a more speech-like profile. Palfreyman and al Khalil's (2003, 20) work on Arabic showed that individuals used local dialect features in IM rather than Modern Standard Arabic. Because Arabic is typically written only in the standard dialect, the use of vernacular lexicon and spellings showed that CMC language mirrors speech rather than writing.

Werry (1996, 61) also argues that CMC "reproduces and simulates the discursive style of face-to-face spoken language." CMC displays informal and speech-like features, including abbreviations, short turn-taking,[5] and omission of auxiliary verbs and pronouns, which he attributes to the "temporal, spatial, and social strictures imposed by IRC [Internet Relay Chat], specifically that messages be typed as quickly and efficiently as possible" (55). Werry's research highlights the importance of the type of medium studied. His study is based on chat-room exchanges, a synchronous group medium. This could explain the difference between his findings and Yates's research, which was based on asynchronous Newsgroup style CMC. Again, assuming similar cross-linguistic effects in CMC, Hentschel's (1998) research provides corroboration of this. In his data, which come from a Serbian chat room, not only were the turns short, but individuals broke sentences into several turns. Hentschel (1998, 10) argues that this strategy kept the attention of the interlocutor while at the same time signaling that there was more coming.

More recent data from American college students suggests that IM is not as quick paced as the chat room environment (Baron 2004, 409). While this slower pace could be an indication of a more formal, written-like register, two points argue against this: IM is a one-to-one form of CMC, and, as

many researchers have discovered, teens multitask while using IM (Nardi, Whittaker, and Bradner 2000; Lenhart, Rainie, and Lewis 2001; Grinter and Palen 2002; Schiano et al. 2002; Baron 2004, 399). This can involve other computer tasks, watching television, doing homework, or having a face-to-face or telephone conversation. This means that the (apparent) slower pace of IM interchanges is likely due to the fact that interlocutors direct less attention to individual conversations, not that they are occupied with careful composition and editing of messages.

Perhaps the most relevant study for our research is Baron (2004), which is one of the few empirical studies of IM at the time of writing. Baron's study is based on a corpus collected from American college students. Her remarkable discovery was that IM was more conservative than is suggested by the press (419). A mere 0.3% of the words had typical IM abbreviations (e.g., *hrs, cuz),* less than 0.8% were initialisms (e.g., *lol, brb*),[6] and 0.4% represented emoticons (smiley faces and the like)[7] (411, 413). Furthermore, where lexical pairs could be contracted, as in *I'm* for 'I am', Baron (2004, 413) found that only 65% were shortened. The study concludes that IM conversations in this corpus "represent a blend of both spoken and written language conventions" (416).

The fact that IM diverges from written language at all raises an alarm from teachers and parents. Indeed, the common perception is that IM, particularly among adolescents, is close to laughable, filled with grammatical errors, incomprehensible words, and secret codes—by all counts, an inferior mode of communication. Teachers often report that student assignments are littered with typical IM abbreviations (Lee 2002). Although it is easy for teachers to pigeonhole CMC generally as the root of this problem, Baron (2003b, 88) offers an alternative explanation: "technology often enhances and reflects rather than precipitating linguistic and social change." Thus, one possibility is that IM is simply mirroring the emerging tendency for written genres to be more like speech, a process referred to as "colloquialization" (Mair 1998; Hundt and Mair 1999; Leech and Smith, 2005). At the same time, the graphic nature of some IM items (e.g., emoticons, elongated spellings, etc.) makes it entirely unlike speech. Both developments reveal novel tendencies, suggesting that IM may actually be a bellwether in the evolution of the English language in general.

DATA AND METHODS

Examination of the literature on CMC reveals that it is still little studied; there are few corpus based studies and only rare comparisons with other forms of communication (e.g., speech, writing), so it has yet to be definitely

characterized. IM, one of the most popular forms of CMC, is even less known. Moreover, the question of precisely where to place it on a spectrum between written and spoken language remains open. This provides us with an intriguing starting point from which to conduct a study of IM. However, embarking on such a study presents a fundamental problem: How does the analyst—typically a middle-aged academic—gain access to the natural informal discourse of the IM speech community, which essentially comprises members of the teenage generation?

As it happens, the Faculty of Arts and Science at the University of Toronto has a unique program that encourages secondary-school students to join university professors in their research projects—called the Mentorship Program.[8] Between 2004 and 2006, the first author (Tagliamonte) set up a research project titled Teen Talk in Toronto that was specifically designed to tap into IM communication. Over the course of the three-year project, two teams of teenagers (Mentees) from high schools within the Toronto District School Board worked on this project. The Mentees became coresearchers and participant observers who engaged in fieldwork and data analysis of an immense corpus of IM language. Table 2 shows the corpus composition by the date of data collection and the birthdate and sex of the participants.

The second author (Denis) aided in the management of these projects as well as the training and supervision of the teenage Mentees. He also collected the data from the 17–20-year-old population from his own social networks in the summer of 2005.

The IM portion of the corpus was constructed from conversational histories—computerized records of the participants' IM interactions with friends. These massive text logs, comprising literally thousands of individual conversations, were donated to the project. An important facet of this IM data

TABLE 2
Corpus Constitution

Birthdate	2004–2005		Summer 2005		2005–2006		Total	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	–	–	–	–	–	2	–	2
1989	–	–	–	–	3	6	3	6
1988	4	4	–	–	2	3	6	7
1987	9	17	3	2	3	2	15	21
1986	–	1	2	4	–	–	2	5
1985	–	–	4	–	–	–	4	–
TOTAL	35 (12)		15 (7)		21 (11)		71 (30)	

NOTE: Numbers in parentheses indicate how many spoken interviews were conducted.

is that the vast majority of the material came from conversational histories that had been saved well before the study began,[9] so the project largely avoids the observer's paradox (Labov 1972, 209)—a notable difference between this and any previous study of IM (e.g., Baron 2004, 404). We believe this may make it one of the most authentic pictures of teen language and interactive CMC discourse in existence and in particular a unique record of IM communication among teenagers.

The Mentees also audio recorded themselves chatting with a subset of the same friends with whom they engaged in IM communication frequently.[10] These conversations were informal and unstructured, modeling the same type of interaction they would have had had the audio-recording device not been present. The data are typically gossip sessions and discussions of school activities and projects.

In sum, our data are a singular documentation of teens interacting with teens. All the conversations are between individuals who were well known to each other, engaged in vibrant, interactive discourse. Moreover, and crucially, the corpora comprise data from the SAME interlocutors recorded in different media: one written (IM) and one spoken. This means we are able to contrast an individual's language use directly between one medium and another.

In total, there are 71 individuals in the corpus, 30 male and 41 female, all born and raised in Canada, all between the ages of 15 and 20 in the years 2004–2006. We emphasize that these corpora have been made possible only by the cooperation and generosity of our Mentees and their friends as well as members of the second author's social networks, all of whom willingly and enthusiastically provided a wealth of data for study.

Following standard ethical guidelines, each individual signed standard consent forms, and all names and screen names were systematically changed in order to ensure anonymity.[11] A formatting protocol was devised in order to process the IM data using Concorder (Rand and Patera 1991). The speech data were transcribed, digitized, and subjected to the same cataloging techniques. All the data were then systematically processed by creating individual concordances and word lists for each individual as well as for the data set as a whole. All told there are over a million words of Instant Messaging and over a quarter of a million words of spoken conversation. To our knowledge, these materials are not directly comparable to any existing body of data.[12]

ANALYSES

In the sections that follow we systematically search and document a range of linguistic phenomena found in the corpora.

IM FORMS. The media typically report IM language by describing a set of characteristic forms with which IM is assumed to be riddled, among them emotional language, such as laughter and other sounds, as in (2), and abbreviations, as in (3).[13]

2. a. [002] HAHAAH what were you on!
   [002] those are great
   [999] I was on 1 hours of sleep :P
   [999] and Blintz
   [DER/002, female, 17; DER/999, male, 20]
   b. [006] thanks darling
   [999] noooooooo problemo :)
   [006] HEHE
   [DER/006, female, 18; DER/999, male, 20]
   c. but now after this horrible editing experience.. i'm back to square one. ARGH. [TEEN/008, female, 16]

3. a. [3] hehe anywhoo i want to go watch tv and eat more advent calender LOL ill TTYL
   [025] LOL ok later
   [TEEN/3, female, 17; TEEN/025, male, 17]
   b. [024] im doing work i leave on wednesday for kenora so i have to get a ton of work done
   [3] kenora for what? basketball? yeah im being buried in HWK
   [TEEN/024, female, 17; TEEN/3, female, 17]
   c. do you have susan peterson's email BTW?
   NVM, clair is getting it [TEEN/027, male, 17]
   d. [999] I'l be down for 7 then and meet you dudes whenever
   [012] K wicked.
   [DER/999, male, 20; DER/012, female, 18]

The first question we pose is how frequent are these forms in IM? Table 3 presents the 16 most numerous of these and their proportion of total word count.

Notice that there are thousands of uses of *haha* and its variants.[14] Indeed, this is the most productive feature by far. The highest frequency form stereotypically associated with IM is *lol*, which, although initially standing for 'laugh out loud', is used by our participants in the flow of conversation as a signal of interlocutor involvement, just as one might say mm-hm in the course of a conversation.[15] The frequency of *lol* is well below *haha*, at just over 4,500 times. Next is *hehe*, another laughter variant but with connotations of giggling.[16] Next comes *omg*, the IM abbreviation for 'oh my god', and then *hmm*, a fairly mundane form expressing contemplation. Thereafter the numbers decrease markedly, and this is where we find many of the famed IM forms,

TABLE 3
Characteristic IM Forms: Frequency and Proportion of Total Word Count

haha (laughing)	16,183	1.47%
lol 'laugh out loud'	4,506	0.41%
hehe (laughing)	2,050	0.19%
omg 'oh my God'	1,261	0.11%
hmm (thinking)	1,038	0.09%
brb 'be right back'	390	0.04%
ttyl 'talk to you later'	298	0.03%
btw 'by the way'	249	0.02%
wtf 'what the fuck'	218	0.02%
arg (frustration)	197	0.02%
hwk 'homework'	99	0.01%
nvm 'nevermind'	78	0.01%
gtg 'got to go'	68	0.01%
np 'no problem'	65	0.01%
lmao 'laugh my ass off'	63	0.01%
nm 'not much'	32	0.00%
TOTAL	26,795	2.44%

NOTE: The data contain many more IM forms, but these are the most frequent.

such as *brb* 'be right back', *ttyl* 'talk to you later', and *wtf* 'what the fuck'.[17] Frequencies of the remaining forms recede into obscurity.

The sheer infrequency of the so-called "characteristic IM forms" is dramatic when we look at the results through accountable proportional analysis. They represent a minuscule proportion of the total number of words in the discourse. Indeed, all the forms combined represent only 2.4% of the corpus. This is remarkably consistent with Baron's (2004) findings for young adults in their early 20s in the early 2000s and a more recent study of college students' use of text messaging on cell phones (Baron and Ling 2003; Ling and Baron 2007).[18] These results confirm—and with a substantially larger and unprompted corpus—that the same is true of the teenage generation in Canada between 2004 and 2006. The use of abbreviations, short forms, and symbolic uses in IM is without a doubt a new vogue, but much rarer than the media have led us to believe.[19]

LOL. Perhaps the most eminent IM abbreviation is *lol* 'laugh out loud'. The frequency of this feature, 4,506 instances in the IM corpus, provides an opportunity to examine the social distribution of laughter variants within this teenage speech community. In particular, since our corpus is comprised of a range of ages, it is possible to test for change over time. Figure 1 shows

FIGURE 1

Distribution of Laughter Variants in IM Data
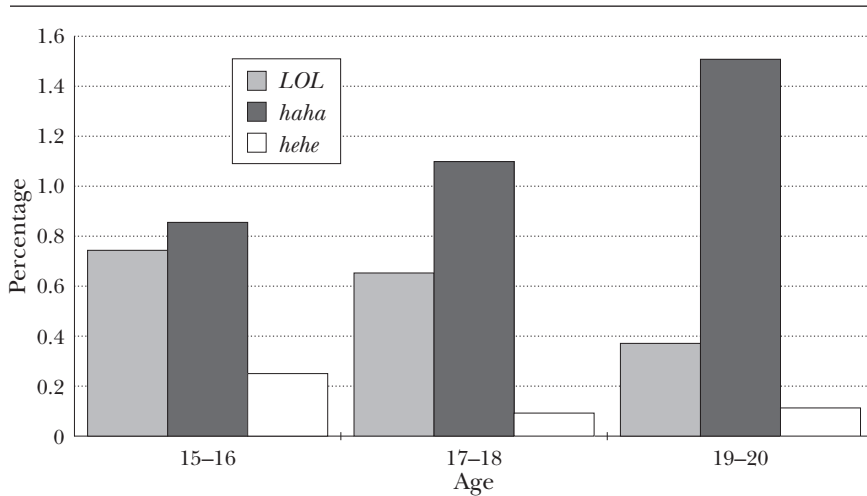
the distribution of the IM laughter variants *lol*, *haha*, and *hehe* according to the age of the individuals.

The use of *lol* declines systematically according to age, with the younger individuals using it the most. In contrast, the more conservative form *haha* is the clear preference of the oldest individuals. Since it is also the case that the older individuals have had the longest exposure to IM, we speculate that this trajectory is the result of incremental loss of the stylized form, *lol*, in favor of *haha*.[20] Perhaps as a result of habituation to the IM environment, it seems that adolescents quickly outgrow at least some of the IM forms. Indeed, a quip from the first author's 16-year-old daughter, "I used to use *lol* when I was a kid," provides anecdotal confirmation of *lol*'s puerile association.

I AND *U*. Another often cited feature of IM is the use of lowercase letters, particularly *i* and *u* for *I* and *you* as in (4).

    4. a. [002] i'm a nerd too!
            [999] yay!
            [002] but whats worse is i'm a hippie nerd
            [DER/002, female, 17; DER/999, male, 20]
      b. i tried getting a hold of u sooo many times [TEEN/024, female, 17]

This variation presents us with a quick and easy way of examining IM language since alternate forms can be tabulated straightforwardly from the text files. Table 4 shows their distribution in the data.

TABLE 4
Distribution of (you) and (I)

you	21,491	91.41%
u	2,020	8.60%
I	10,581	25.93%
i	30,222	74.07%

Orthographic *you* represents the vast majority of variable (you) forms—91%. Contrast this with variable (I), where the lowercase variant reigns at 74%. Thus, use of *i* stands out as a frequent feature of IM overall. Of course the choice of an abbreviated form (which applies only to *u*) and use of upper- versus lowercase (which is fundamental to the use of lowercase *i*) are different processes. This may explain the contrasting findings. However, the next step is to consider whether or not these features are used similarly across individuals.

Figures 2 and 3 show the distribution of variable (you) and (I) by individual. Figure 2 shows that by far the majority of individuals use the standard form, orthographic *you* categorically. There are only a few individuals who use lowercase *u*, and of these most use it the majority of the time. Interestingly, very few individuals are variable. Figure 3 shows that by far the majority of individuals use the nonstandard form lowercase *i*. Only some use the standard form *I*. The norm for sociolinguistic variables is that individuals will use competing forms (Sankoff 1980, 55). Here, however, only a few individuals do. Thus, the use of neither lowercase *u* nor lowercase *i* exhibits the profile one would expect of a typical linguistic variable. Instead, it appears that the selection of one variant or the other is an individual's stylistic choice, with

FIGURE 2
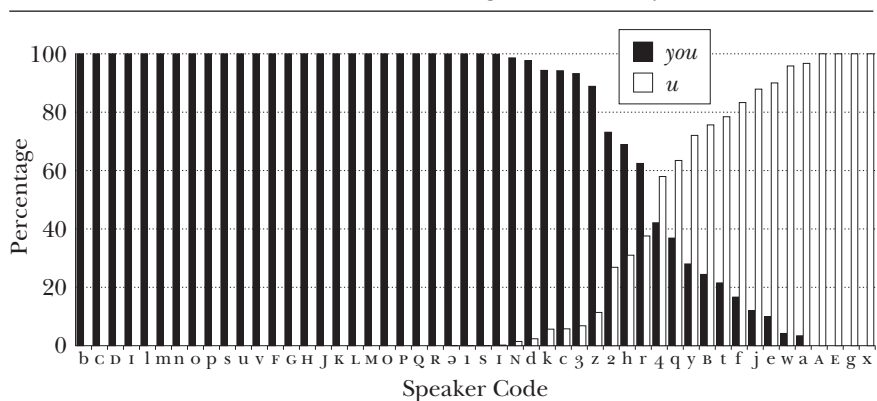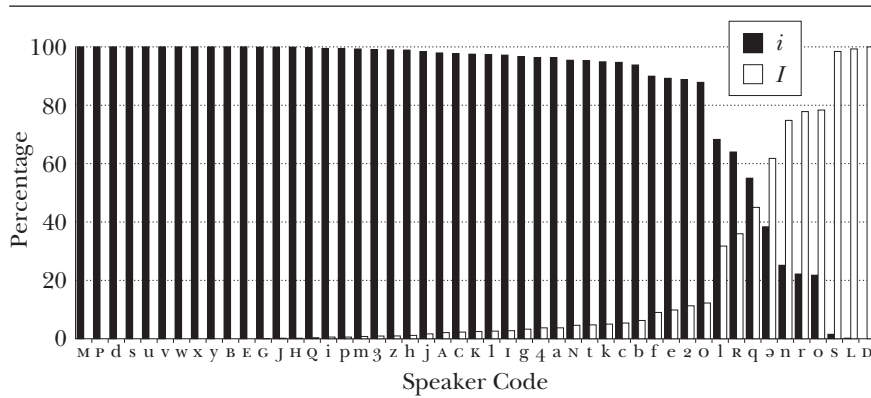Distribution of Second-Person Singular Pronouns *you* and *u*

FIGURE 3
Distribution of First-Person Singular Pronouns *I* and *i*

most individuals opting for one form or another.[21] Indeed, previous studies suggest that these IM forms can have strong in-group affiliations (Cherny 1999, 92). This is another result pointing to the stylized nature of the so-called IM features; however, we still do not know much about the nature of the IM linguistic system. We begin this inquiry by targeting a feature thought to be a good measure of the spoken-to-written continuum.

PERSONAL PRONOUNS. According to earlier research, first- and second-person pronouns are more frequent in speech than in writing. In order to situate CMC on this spectrum, Yates (1996) examined the distributions of pronouns of personal reference. He found that the language used in computer conferencing was much more like speech than writing.

Taking Yates's lead, we extracted all pronouns of personal reference in both the IM and speech data. Figure 4 presents these results proportionally and compares them to Yates's written data. In the speech data, personal reference is dominated by first-person pronouns at 53%. The IM data exhibits a very similar pattern—first-person pronouns appear 62% of the time. These findings are parallel to Yates's findings. Yet, comparison of the IM data with Yates's written data reveals a stark contrast. In writing, third-person pronouns dominate. This supports the hypothesis that IM is similar to speech and different from writing. Unlike the choice between *i* and *u,* the distribution of personal pronouns reflects the CONTENT of the discourse, such as topic of discussion, suggesting that IM's register/genre attributes are speechlike in nature.

Thus far, the emerging profile of IM language reveals a stylized medium containing a relatively small set of identifying features but which is more

FIGURE 4
Distribution of Personal Pronouns by Medium
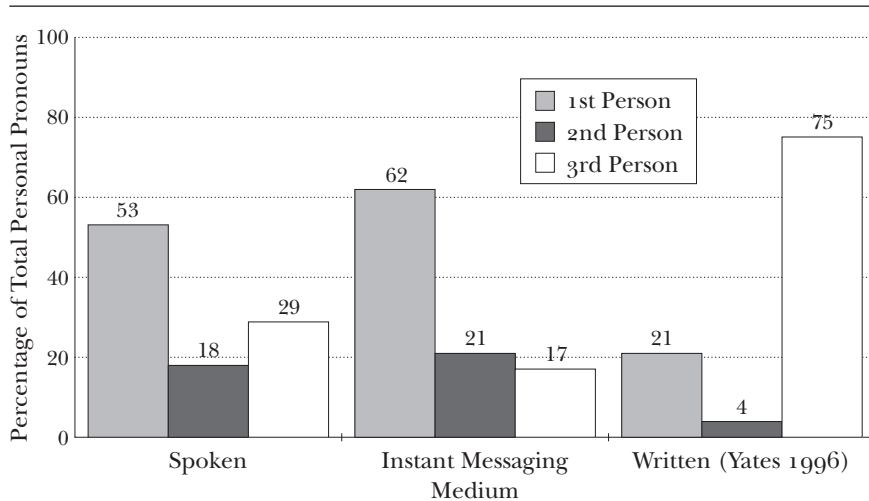
conservative than the popular press has suggested. Yet patterns of linguistic usage, as demonstrated by pronominal choice, suggest a speechlike register. It is at this point that we launch into a series of quantitative variationist analyses of different areas of English grammar. Use of variable grammatical features is thought to reveal deeper, systemic aspects of the linguistic system (Poplack and Tagliamonte 2001, 93). As such, they permit the analyst to tap into the underlying grammar. We have selected these particular features for two reasons. First, they are features that are undergoing change in the English language more generally, thus enabling us to assess where IM is positioned vis-à-vis ongoing linguistic change. Second, each of the variables has been studied extensively. In addition to having access to the results of earlier studies, we also mimic these previously established methods and analysis, making the results obtained here comparable across studies. As far as we are aware, such an analysis of IM has not yet been conducted. An added benefit of this corpus is that it allows for cross-genre comparison of the same teenagers in near identical contexts, but in different media.[22] This provides a window into the nature of IM. If it is more like spoken language, then we can expect it to pattern similarly to the speech data.[23] If, however, IM is on the cutting edge of innovation, we can expect it to have greater use of incoming forms than speech. Furthermore, because the grammatical features we target for investigation involve a combination of formal, standard, and colloquial variants, the comparative distribution of variant forms in the IM and speech corpora will provide a unique opportunity to document the sociolinguistic nature of IM language.

INTENSIFIERS. The intensifier system in Toronto English is undergoing rapid change, particularly among adolescents (Tagliamonte forthcoming). Intensifiers are adverbs that maximize or boost meaning, as in (5).

5. a. my clean room is so weird [TEEN/004, male, 18, IM]
    b. haha it was kinda creepy, but VERY cool. [TEEN/002, female, 17, IM]
   c. I don't have any stories though my life is REALLY boring. [DER/003, male, 19, speech]

This area of grammar is well known to change continuously (Stoffel 1901, 2), partly because of a "speaker's desire to be 'original,' to demonstrate his or her verbal skills, and to capture the attention of their audience" (Peters 1994, 271). This presents us with an ideal feature to investigate in the IM medium and to test the common hypothesis that IM is on the leading edge of linguistic innovation (Schiano et al. 2002, 1). In earlier research on spoken corpora of British English, Ito and Tagliamonte (2003, 257) discovered that *very*, the most frequent intensifier in contemporary English usage (Fries 1940, 201; Bäcklund 1973, 290), is frequent only among older individuals. In contrast, the newer variant *really* is increasing among the youngest generation. In Toronto English, the same variety spoken by the teenagers in this study, younger individuals are shifting dramatically toward use of *really*. Moreover, a new form is on the horizon, intensifier *so* (Tagliamonte forthcoming).

 In order to study intensifiers in the speech and IM corpora, we adapted the methodology in Ito and Tagliamonte (2003), including in the analysis the 25 most commonly used adjectives from Tagliamonte's (2003–2006, forthcoming) corpus of Toronto English.[24] Following standard variationist practice and the principle of accountability (Labov 1972, 72), we extracted all of these adjectives in contexts in which they were capable of being intensified, whether they were or not. This amounted to over 16,320 tokens across the two data sets.[25]

The questions are: What intensifiers are used in IM? Where will IM be positioned vis-à-vis speech? If IM is a speechlike medium, then we can expect parallelism with the speech data. Furthermore, assuming IM is on the leading edge of change, we might expect newer forms such as *really* and, in particular, *so* to be more frequent as well.

Figure 5 shows the distribution of the main intensifiers represented as the proportion each intensifier represents of all the adjectives considered. The most frequent forms in these data, both speech and IM, parallel those found in spoken data in earlier research—they are *really*, *so*, and *very*.[26] Consistent with earlier research, *very* is rare among adolescents. However, it is the relative proportion of *really* and *so* that are informative. In our speech data, *really* and *so* are nearly equally used for intensifying adverbs. In contrast, in IM the teenagers are much less likely to use intensifiers at all. Since the use

FIGURE 5
Distribution of Major Intensifiers by Medium
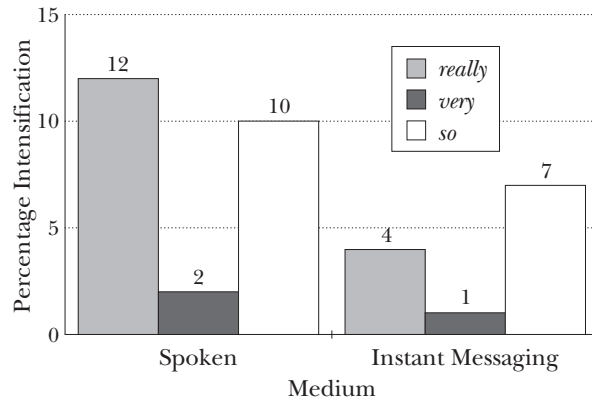
NOTE: Percentages were calculated based on the total number of intensifiable adjectives. As is typical, most were not intensified, which explains the low percentages.

of intensifiers is typically associated with colloquial usage and nonstandard varieties (e.g., Stoffel 1901; Fries 1940), this suggests that IM is a more formal register than speech. However, when the teenagers do intensify their adjectives, their preferred choice is *so.* Thus, while the two media mirror each other in exhibiting the same intensifying adverbs overall, there are some notable differences. This is the first indication that IM cannot simply be labeled a speechlike register. It mixes innovative trends alongside an overarching conservative nature.

THE QUOTATIVE SYSTEM. We now turn to a feature of English that is another excellent test for language change in action—the quotative system. In this area of grammar, the vernacular variant *be like* is strongly associated with young people and has increased exponentially in the last decade (e.g., Tagliamonte and D'Arcy, 2004a, 2004b, 2007a). Quotative *be like*, similar to intensifier *so*, provides a choice opportunity to find out how a rapidly innovating feature operates in IM.

The examples in (6) show typical usage of quotative verbs in the corpora.

6. a. when I ASKED him, "So do you still like Anastasia?"
        … Ø "Well, of course she 's good-looking,"
        … and I WAS LIKE, "Woah! Not what I asked!" [TEEN/8, female, 16, speech]

  b. He's LIKE "come on come on"
   Ø "pow"
   Ø "oh god."
   He's LIKE– he's LIKE, "if we fought anyone, it would be The-Antics
    [DER/001, male, 19, IM]
  c. I e-mailed him, I SAID, "I despise skirts, can I please wear dress pants?"
    [TEEN/039, female, 16, speech]
  d. some guy's running around the street GOING "World War Three, oh my
    god, it 's world war three!" Ah, do you remember that? [DER/001,
    male, 19, speech]
  e. [003] we were taking to PL, and she WAS LIKE … you guys will need to
    explain why your error was so huge
   [003] and he WAS LIKE "uh … yeah…."
   [003] I don't know. People at uts can be so fucki
   [1] yeah
   [TEEN/003, female, 16; TEEN/1, female, 16; IM]

   In earlier research on spoken corpora, the incoming form *be like* was found to be encroaching on all the other forms in the contemporary English quotative system, particularly in the younger generation and especially among young women (e.g., Tagliamonte and Hudson 1999; Buchstaller 2001a, 2001b, 2004; Singler 2001; Cukor-Avila 2002; Tagliamonte and D'Arcy 2004a, 2007a). In contrast, the standard variants, *say*, *think*, and the like are the favored forms among older individuals.

   For this analysis, we extracted the first 50 quotative verbs from individuals for whom we had both IM and speech data. This amounted to nearly 500 tokens. Then we followed the methodology for coding and analysis outlined in Tagliamonte and D'Arcy (2004a).

   The questions are: What quotatives are used in IM? Where will IM be positioned vis-à-vis speech? If IM is speechlike, then we can expect a greater use of nonstandard forms, especially *be like.* Furthermore, assuming IM is on the leading edge of change, we might expect *be like* to be used more frequently than in speech. Figure 6 presents the distributions of the three most frequent quotatives in the data—*be like*, *said*, and zero.

   There is a dramatic difference between IM and speech media. The spoken materials reflect the profile of teen language we would expect. Consistent with many other studies, *be like* is by far the dominant form, at 68%. However, the unexpected result is how conservative IM is in comparison. Although it is true that there is not nearly the same frequency of quotative verbs in the IM data as in the speech data,[27] the fact that we have used proportional analysis provides an opportunity to compare the two media accountably. Zero forms are dramatically more frequent in IM (41%) than in speech (12%).

figure 6
Distribution of Quotative Verbs by Medium
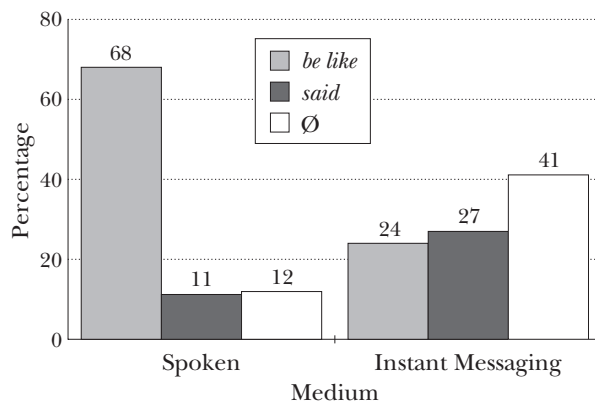
This is undoubtedly due to the fact that alternating dialogue can be marked clearly by separate transmissions and/or punctuation, as in (6b), suggesting that the heightened use of zero in IM is a genre-specific effect. However, this cannot explain the curtailed use of *be like* in IM as compared to speech. In the speech data, the use of *be like* dominates the system, whereas in IM the formal variant *said* and *be like* are robust competitors. This difference is all the more remarkable since quotative *said* is rarely used by the teenagers in their speech. Yet in IM the very same teenagers use it just as much as *be like*.[28]

Furthermore, the IM data comprises an entirely different balance in the inventory of quotatives: *be like, said*, and zero appear in proportions of 24%, 27%, and 41%, respectively. In contrast, the speech data is characterized by overwhelming use of a single form—*be like*, which supports the conclusion that IM has a wider range of variants and greater use of formal and standard variants than spoken language.

future temporal reference. What if we go deeper into the grammar and examine a feature that has been variable in English for a very long time? The future temporal reference system of English is in the process of longitudinal change in which the incoming form *going to* has gradually been increasing since the 1400s (see, e.g., Poplack and Tagliamonte 1999). In this change, none of the variants are particularly stigmatized; however, the *go* future has connotations of informality generally, and its reduced variants, including *gonna*, can be considered vernacular.

The examples in (7) show that the IM data has the gamut of forms attested in contemporary English—*going to, gonna, will, 'll, shall*, and the simple or periphrastic present as in (7d).

7. a. star wars episode III is GOING TO suck just as much as the previous two and everyone is GOING TO be really sad [DER/004, male, 18, IM]

b. by the way there WILL be a pre drink/jam session at my house new years eve [DER/010, male, 16, IM]

c. hmm....11:30...i think i SHALL hit the sack tonight ... [TEEN/050. male, 19, IM]

d. and, as far as i know, i'm going shopping with her *alone* on sunday [TEEN/034, male, 16, IM]

e. ILL get my shift switched! nick too? how're u GONNA fit us all in, lol. [TEEN/003, female, 16, IM]

We adapted the methodology deployed in earlier research (Poplack and Tagliamonte 1999) but included only future-in-the-present contexts.[29] This amounted to over 4,500 tokens across the two data sets.

The questions are: What future variants are used in IM? Where will IM be positioned vis-à-vis speech? If IM is more speechlike, then we can expect parallelism with the speech data. Furthermore, assuming IM is leading change, we might expect newer, incoming forms such as *going to* to be more frequent. Figure 7 shows the distribution of future variants in the data.

Overall, the two media pattern in tandem. *Will* is the most frequent form. The two variants of the *go* future, *going to* and *gonna*, make up the bulk of the remainder, while the use of the simple and periphrastic present is low and *shall* is virtually nonexistent. As with intensifiers and quotative verbs, these relative frequencies parallel the findings of earlier research. As far as differences between the two media are concerned, the standard variant *will* is used more in IM (56%) than in speech (46%), while *going to/gonna* is used

FIGURE 7
Distribution of Future Temporal Reference Variants by Medium
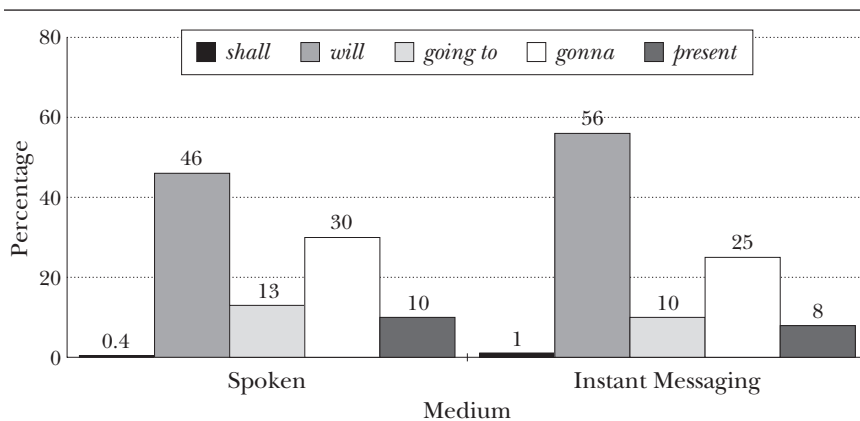
less in IM (35%) than in speech (43%).[30] Focusing on the *go* future alone, this trend is repeated. The more vernacular variant *gonna* is more frequent in speech (30%) than in IM (25%). In sum, when these teenagers use IM, they tend more toward the standard variants than when they are talking to each other. IM is, once again, more conservative than speech.

modals of necessity. The last feature we will consider is also a feature of English undergoing change. This area of grammar is the deontic modality system, which involves the expression of obligation and necessity, as in (8).

8. a. As soon as this is over I gotta go up to Forrest-Hill, grab Audrey, then we
       have to go up to Yorkdale. [DER/999, male, 20, speech]
   b. [999] I wonder what my maxim magazine says about a girl inviting a guy
       to watch a movie…hmmm
      [999] I must consult my Man-Bible
      [002] report back
      [DER/999, male, 19; DER/002, female, 17; speech]
   d. You've gotta send me the pics. [TEEN/8, female, 16, IM]
   e. [2] bought the dvd, came out YESTERDAY
      [015] OH OKAY, I NEED TO GET THAT
   [TEEN/2, female, 17; TEEN/015, female, 17; IM]

In earlier research on Toronto English, Tagliamonte and D'Arcy (2007b) demonstrated that teenagers use *have to* most of the time. The other variants of deontic modality appear infrequently and, with the exception of *need to*, which is a minority variant across people of all ages, are mostly restricted to the older individuals. As with the increasing use of the *go* future, the change is below the level of awareness and, unlike intensifying adverbs, the incoming form *have to* is not overtly associated with teenagers.

Once again our methodology replicates that of earlier research, in this case Tagliamonte and D'Arcy (2007b), who limit the analysis to tokens of deontic modality that are variable between *must, have to, have got to, got to,* and *need to* (including all morphological forms, such as *have to, has to, and needs to*). This provided over 2,200 contexts.

The questions relevant for this analysis are: How does IM reflect ongoing change in deontic modality in Toronto English? Where will IM be positioned vis-à-vis speech? If IM is more speechlike, we can expect parallelism with the speech data. Furthermore, assuming IM is on the leading edge of change, we might expect the incoming form *have to* to be more frequent than in speech. Figure 8 shows the distribution of variants of deontic modality.

As might be expected, *have to* dominates. Notice, however, that its frequency is highest in the speech data, at 83%. Furthermore, in the speech data,

FIGURE 8
Distribution of Deontic Modals by Medium
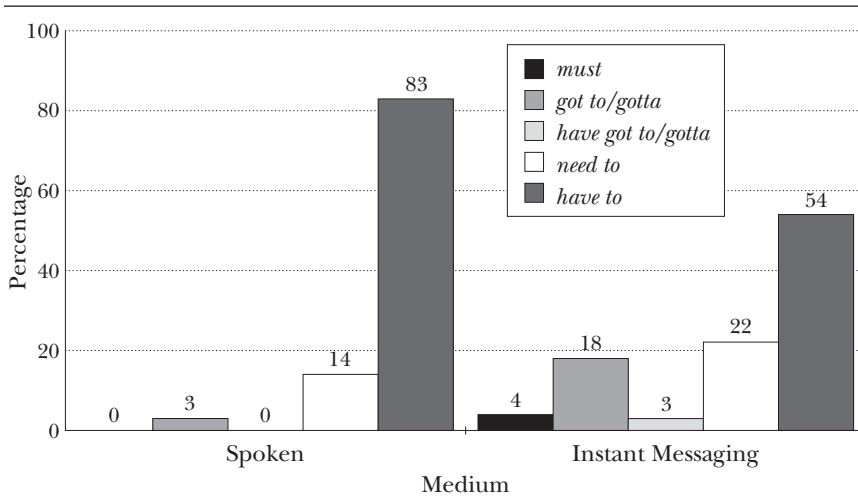
neither *must* nor *have got to/have gotta* nor *got to/gotta* represent more than a few tokens apiece. Only the variant *need to*—the feature that is stable across both time and space in Toronto—is found with any degree of frequency, at 14%. Moreover, this rate approximates the rates reported in Tagliamonte and D'Arcy (2007a).

This time the comparison between IM and speech uncovers two important results that build on what we have already found. First, observe the comparatively infrequent use in IM of the variant that dominates in spoken language: compare the proportion of *have to* in speech (83%) and that in IM (54%). Second, IM shows the full complement of all the variants present in this system in the language—formal, (i.e., *must*), standard (i.e., *have got to*), and colloquial (i.e., *got to*). The common theme is the unique combination of variants in the IM medium.

SUMMARY. The overall findings are summarized in table 5, which reveals a number of striking convergences in the nature of IM language. Contrary to expectation, speech has a more innovative profile in comparison to IM. The incoming forms *so, be like*, and *have to* are all more frequent in the speech data. At the same time, speech tends to be more vernacular, containing higher rates of *going to* than *will*. In contrast, IM language is consistently more varied, exhibiting a wider range of variants than speech and, in particular, containing a higher proportion of standard forms than speech.

TABLE 5
Summary of Findings

Analysis	Results
CMC Features	
acronyms, short forms, emotional language	use of these forms is rarer than common perception
lol	systematic decline of *lol* according to age suggests such forms are outgrown
you/u and *I/i*	use of one form or the other is an individual stylistic choice
Inherent Variability	
intensifiers	less overall intensification in IM; in SP *really* and *so* predominate, but in IM innovative *so* is the leading form
quotatives	innovative form *be like* dominates in speech; more robust use of varied forms in IM including a heightened use of the stardard variant *said*
future temporal reference	vernacular *gonna* more frequent in speech; a wider range of variants in IM, more frequent use of standard *will*
modals of necessity	incoming form *have to* more frequent and dominant in speech; wider range of variants in IM than in speech (including formal *must* and vernacular *gotta*)

CONCLUSIONS

Taken together these findings permit us to make a number of observations. In a million and a half words of IM discourse among 71 teenagers, the use of short forms, abbreviations, and emotional language is infinitesimally small, less than 3% of the data. Indeed, the proportion of use by these adolescents is nearly identical to the rates found for the same features among 20-year-olds (Baron 2004). These results challenge the adverse perceptions of IM promulgated in the media and suggest that they have been overblown. Indeed, consistent with Baron's (2004) suggestion that the college students in her study had outgrown the stylized IM forms (quoted in Ghenu 2005, 8) the teenagers in our study are already conservative. Indeed, if our results are indicative, adolescents appear to abandon the "funky" IM features at a very young age. Nevertheless, we have clearly tapped into a vibrant new medium of communication with its own unique style (see also Ferrara, Brunner, and Whittemore 1991; Crystal 2001; Herring 2003, 2004). Moreover, our results corroborate earlier CMC research in demonstrating that language use in IM

is part of a much broader contemporary trend toward more informal language generally (e.g., Biber 1988; Biber et al. 1999; Hundt and Mair 1999). Yet, the variety of English used in the IM corpora we have studied here is neither a caricature of real language nor some kind of basilectal lowlife. But the question is—what is it?

Our foray into the IM environment through quantitative sociolinguistic analysis, encompassing four areas of grammar and over 20,000 individual examples, reveals that IM is firmly rooted in the model of the extant language, reflecting the same structured heterogeneity (variation) and the same dynamic, ongoing processes of linguistic change that are currently under way in the speech community in which the teenagers live. For each of the linguistic variables we studied, the IM materials mirror the larger Toronto community in terms of the major variants in the system. Yet the register we have tapped into is "a language variety that never existed before" (Ferrara, Brunner, and Whittemore 1991, 26). The consensus in the literature is that IM is a hybrid. The findings we have presented here permit us to document just what kind of species it is. For every linguistic variable, IM demonstrates a unique fusion of variants. Simultaneously, it makes use of formal variants such as *shall* and *must*; informal variants such as *will* and *have to*; and highly colloquial variants like *gonna* and *gotta*. It also contains heightened use of new innovations such as intensifier *so*, providing evidence to confirm that IM is a medium on the forefront of change.

This blend of features is easily visible in many of the examples throughout the article, but the following excerpt from a single conversation illustrates it particularly well. In this interchange, the two male interlocutors are engaged in a discussion about their new band. They are both excited about the possibility of playing together and writing new songs. The conversation revolves around arrangements having to do with renting recording equipment and getting together.[31]

9. [001] ye guy were gonna do lotta different shit for sure
   [001] thesongs are GONNA totally range
   [999] for sure
   [001] so you come out tonight
   [999] i HAVE to see, but i WILL try, i wanna
   [001] ok just us 3
   [001] were GONNA party and write shit
   [999] nice
   [001] ye guy we went to long and mcquade the all in one
   [001] is soooooooo cheap
   [001] we can rent it for 2 monthes for 240 bux
   [001] all in one mixer
   [999] really? that's wicked

[001] yep so like rest assured all our band money WILL be able to cover record-
ings and me and dick GOTTA talk to you about recordings, as you already
know but were GONNA see how it goes at anti warped and see if it should
just be us 3 doing the recordings and what not, i dunno its something
dick was goin on about but yes meet tonight i MUST run and shower be
back later

[DER/001, male, 19; DER/999, male, 19; IM]

There are numerous typos, misspellings, swear words, and colloquialisms
(i.e., slang), along with numeric forms and a propensity toward lowercase;
however, notice that there are no IM forms at all. Moreover, a closer exami-
nation of the data reveals the type of linguistic mixing that we have seen re-
peated feature after feature in the quantitative analyses. We have highlighted
the forms that have been investigated here, although there are many others
that are clearly worthy of study. It is important to point out that the type of
combination of features that the quantitative analyses have revealed is not
due to amalgamating different conversations from different registers. Even a
single turn may contain variants of contrastive formality, as with *have to, will,*
and *wanna* in line five. Note especially in the lengthy last turn by individual
[001] that the standard variants *will* and formal *must* appear alongside *gotta*
and *gonna.* Moreover, among innumerable other nonstandard and colloquial
features, the phrase structure that underlies the discourse is actually quite
complex, including subordinators typical of written registers, such as *but* and
if and even a highly formal use of *as.*

We have selected the following vibrant examples to illustrate contras-
tive formality in the same turn. The use of *shall* occurs alongside *serious* as
an intensifier and the slang term *jam* in (10a). The use of quotative *says* in
(10b) occurs where we would certainly expect *be like* in speech. The use of
deontic *must* in (10c) appears with the formal variant *very* in (10d). In addi-
tion, these particular examples also include three IM abbreviations, *omgod,*
lol, and *ttyl.*

10. a. aaaaaaaaagh the show tonight SHALL rock some serious jam [DER/001,
male, 19, IM]
    b. Jeff SAYS "lyk omgod omgod omgodzzzzzZZZzzzzz!!!11one" [TEEN/8,
female, 15, IM]
    c. heheh okieee! MUST finish it now ill ttyl [TEEN/003, female, 17, IM]
    d. lol.. as u can tell im VERY bitter right now. [TEEN/008, female, 18,
IM]

This consistent juxtaposition of "forms of a different feather" is the quintes-
sential characteristic of IM discourse.

What would lead to such a development? In formal written language, colloquial variants are illegitimate, prohibited by "language police" of all persuasions (teachers, editors, etc.). Indeed, anecdotal reports suggest that teachers are increasingly penalizing students for the use of abbreviations in written assignments. In teenage conversations, however, formal variants are equally undesirable. IM appears to be a venue in which teenagers are free to use all these features together. This linguistic fusion is endemic to the register itself. Individuals pick and choose from all the available variants that their linguistic system has to offer and draw from the entire stylistic repertoire of the language that exists at a given point in time. If the teenagers did not already possess skilled command of their linguistic system, this would be impossible. The character and nature of IM we have uncovered here reveals fluid mastery of the sociolinguistic resources in their speech community. We conclude that IM, and perhaps computer-mediated communication more generally, is not the ruin of this generation at all, but an expansive new linguistic renaissance.

Indeed, this study of IM language is likely already behind the times and taps only a very small part of what is even now developing. The insurgence of new media into the contemporary world of communication is always expanding. Since we conducted this study, newer and trendier ways to interact online have developed, including mushrooming social-networking Web sites (e.g., MySpace, Facebook) (see Baron et al. 2005), multiplayer online role-playing games (e.g. World of Warcraft), and undoubtedly untold other newfangled ways to communicate online. Simultaneously, the use of text messaging on mobile phones has gained in popularity in North America, providing yet another medium that will shape and reshape the nature of communication (see Baron and Ling 2003; Ling and Baron 2007). All these provide yet to be discovered venues in which the foremost commodity is language. To seek out and study the "intriguingly new and still evolving linguistic varieties" (Crystal 2006, 271) that will emerge from these developments will be a gateway to understanding the future of human communication and perhaps even greater insights into the language faculty itself.

NOTES

1. We preserve the IM material precisely as it was typed, complete with typos, spelling errors, spacing anomalies, and the like.

2. The referencing system for our examples records the corpus and individual. "DER" indicates that the example was collected by Derek Denis, the second author. "TEEN" indicates that the example was collected by the Mentees. The numbers following the corpus designation indentify each individual: three digits indicate the interviewee; single or double digits indicate the interviewer.

3. "Structured heterogeneity" is the idea that variation is inherent to language and further that it is neither random nor free, but systematic and rule governed. The question then becomes how these systemic patterns are reflected across different media.

4. See Baron (2003a) for an in-depth discussion on the history and usage of different types of CMC.

5. "Turn" in synchronous CMC is the term for an individual message of any length sent by a user.

6. Baron (2004, 411) excluded abbreviations and initialisms that are common in offline usage (e.g., *hrs* 'hours') or are direct representations of spoken usage (e.g., *cuz* 'because').

7. Emoticons are among the most novel features of IM. They are "abbreviations of expressions of mood, tone of voice, or instruction to the reader" (Randall 2002, 27) made using typographical symbols (e.g., :-) 'smile', ;-) 'wink', :-P 'stick out one's tongue').

8. For information about the University of Toronto's Mentorship Program, see http://www.artsci.utoronto.ca/prospective/highschoolprograms/enrichment#1.

9. Note that the MSN software allows users the option of saving a text record of their conversations. It was our luck that the Mentees in our study had previously saved documents of this type on their computers when our project began. Many of them simply opted to donate these files to the project.

10. Due to the amount of IM data that the Mentees provided, the overall corpus is more heavily weighted toward IM than spoken communication.

11. Individual conversations were not tabulated separately. All parties involved were given the option to remove any sensitive interchanges; however, such excisement

must have been rare since the data often contain extremely personal interactions.

12. The closest comparison is the corpus collected by Yates (1996), which comprises written materials from the Oslo-Lancaster Corpus, the London-Lund Corpus of Spoken British English. The CMC data come from computer conferencing data at the Open University in the United Kingdom—mostly communications between students and professors about course work (Yates 1996, 30–33).

13. We did not consider emoticons in this study. The versions of the program at the time of our study automatically converted the keystrokes that make up emoticons into a graphical representation. These graphics, however, do not translate across platforms straightforwardly, making it difficult to study them systematically using the methods we have employed here.

14. This tabulation comprises all variants of *haha*, including *hahaha*, *hah*, *ahahaha*, and the like.

15. The use of *lol* as a signal for interlocutor involvement is consistent with Baron's (2004, 416) comment that *lol* functions as a "phatic filler comparable to *okay*, *really*, or *yeah* in spoken discourse."

16. This tabulation comprised all variants of *hehe*, including *hehehe*, *heh*, *eheheh*, and the like.

17. Note that earlier CMC abbreviations reported in Baron (2003a, 21) and Crystal (2001, 85–86) seemed to our participants to be old-fashioned. This may well be why they are not used in the current corpus. Moreover, according to anecdotal reports from our participants, many of the frequently attested IM forms in the literature were unknown to them at the time of writing. For example, one of our participants had been using IM for 4–5 years but had never seen *lmao*. As one of our participant-observers reported: "it's not that she's uncool but rather *lmao* is." Such observations, however anecdotal, point to the transitory nature of IM abbreviations and acronyms generally.

18. Baron (2004) analyzed nine conversations between females; nine between males; and five between male and female. The total number of words in her corpus was 11,718.

19. Although people use abbreviations all the time in speech (e.g., *TV*, *DVD*, *US*, *UK*), to our knowledge no one has yet conducted an accountable study of abbreviations or short forms in regular use in speech or writing or how such usages have changed over time.

20. Since *lol* is among the most well established IM forms (e.g., Cherny 1999, 92), it seems unlikely that the heightened use of *lol* among the younger individuals is the result of increasing frequency due to innovative change.

21. One might think it informative to examine the individuals who are variable to determine whether they exhibit systematic patterning of these alternate forms or not. Preliminary analysis revealed no obvious factors that could distinguish these individuals. They were both male and female, older and younger.

22. This is a major strength of the present study. It studies data from Instant Messaging and spoken conversations from (mostly) the same individuals, whereas

earlier research comparing CMC and speech rely on different data sources and different individuals.

23. A further test would be to accumulate written materials from the same speakers. Future research should include such data.

24. These 25 most commonly used adjectives represent over 75% of all adjectives from the Toronto English Corpus, thus we extrapolated from this that they would comprise the majority of the adjectives in these IM and speech data as well.

25. Extraction was greatly facilitated by running the huge data files through Concorder (Rand and Patera 1991), selecting the relevant forms, and then importing the text files comprising them directly into Goldvarb X (Sankoff, Tagliamonte, and Smith 2005).

26. In this analysis, as well as the subsequent ones, we will not go into the details of variant distributions by age and sex. The few contrasts between male and female speakers did not detract from the major trends we will report and in all cases followed expected sociolinguistic trends (e.g., females had higher frequencies of standard variants). A more in-depth consideration of social factors is left for future research.

27. There are 6.3 quotatives per thousand words in the speech data and only 1.7 per thousand words in the IM data.

28. While both intensifier *so* and quotative *be like* are incoming forms, why is it that only *so* is innovative in IM? The answer undoubtedly lies in the contrasting trajectories of these two changes. While *be like* has been developing for several generations and is used by individuals right up to the late 30s (Tagliamonte and D'Arcy 2007a), the increasing frequency of intensifier *so* is restricted to adolescents. Its innovative status in IM is consistent with this profile.

29. This circumvents methodological problems arising from defining the variable context for future-in-the-past contexts (see Poplack and Tagliamonte 1999).

30. One might alternatively hypothesize that, because *will* is a shorter form than *going to* or *gonna*, it is predisposed to occur in the IM environment. The same argument could be made in the case of intensifier *so*, which is the shortest of the intensifiers. However, we are reticent to propose this explanation. Having heard this suggestion numerous times, we analyzed the two corpora for average word length (not shown here) and found that the IM data actually have a longer average word length score than the speech data, 3.87 vs. 3.78 letters per word.

31. Several lexical items require explanation: "long and mcquade" is a Canadian music store; "all in one" refers to a type of recording device; "anti warped" is the name of a concert where the band was to play; and "bux" is a slang term for the dollar.

REFERENCES

Axtman, Kris. 2002. "'r u online?': The Evolving Lexicon of Wired Teens." *Christian Science Monitor*, Dec. 12, 1. Available at http://www.csmonitor.com/2002/1212/p01s01-ussc.html.

Bäcklund, Ulf. 1973. *The Collocation of Adverbs of Degree in English*. Studia Anglistica Upsaliensia 13. Uppsala, Sweden: Acta Universitatis Upsaliensis.

Baron, Naomi S. 2003a. "Language of the Internet." In *The Stanford Handbook for Language Engineers*, ed. Ali Farghali, 1–63. Stanford, Calif.: CSLI Publications,.

———. 2003b. "Why Email Looks Like Speech: Proofreading Pedagogy and Public Face." In *New Media Language*, ed. Jean Aitchison and Diana M. Lewis, 85–94. London: Routledge.

———. 2004. "See You Online: Gender Issues in College Student Use of Instant Messaging." *Journal of Language and Social Psychology* 23: 397–423.

Baron, Naomi S., and Rich Ling. 2003. "IM and SMS: A Linguistic Comparison." Paper presented at the Fourth International Conference of the Association of Internet Researchers, Toronto, Oct. 16–19. Available at http://jls.sagepub .com/cgi/reprint/23/4/397.pdf.

———. 2007. "Emerging Patterns of American Mobile Phone Use: Electronically-Mediated Communication in Transition." In *Mobile Media 2007: Proceedings of an International Conference on Social and Cultural Aspects of Mobile Phones, Convergent Media and Wireless Technologies*, ed. Gerard Goggin and Larissa Hjorth, 218–30. Sydney: University of Sydney.

Baron, Naomi S., Lauren Squires, Sara Tench, and Marshall Thompson. 2005. "Tethered or Mobile? Use of Away Messages in Instant Messaging by American College Students." In *Mobile Communications: Re-Negotiation of the Social Sphere*, ed. Rich Ling and Per E. Pederson, 293–311. London: Springer.

Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge Univ. Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, England: Longman.

Buchstaller, Isabelle. 2001a. "An Alternative View of *like*: Its Grammaticalisation in Conversational American English and Beyond." *Edinburgh Working Papers in Applied Linguistics* 11: 21–41.

———. 2001b. "*He Goes* and *I'm Like*: The New Quotatives Re-visited." Paper presented at the 30th annual meeting on New Ways of Analyzing Variation (NWAV 30), Raleigh, N.C., Oct. 11–14. Available at http://www.ling.ed.ac.uk/~pgc/ archive/2002/proc02/buchstaller02.pdf.

———. 2004. "The Sociolinguistic Constraints on the Quotative System: British English and US English Compared." Ph.D. diss., Univ. of Edinburgh.

Chafe, Wallace. 1982. "Integration and Involvement in Speaking, Writing, and Oral Literature." In *Spoken and Written Language: Exploring Orality and Literacy*, ed. Deborah Tannen, 35–53. Norwood, N.J.: Ablex.

Cherny, Lynn. 1999. *Conversation and Community: Chat in a Virtual World*. Stanford, Calif.: CSLI Publications.

Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge Univ. Press.

———. 2006. *Language and the Internet*. 2nd ed. Cambridge: Cambridge Univ. Press.

Cukor-Avila, Patricia. 2002. "*She say, She go, She be like*: Verbs of Quotation over Time in African American Vernacular English." *American Speech* 77: 3–31.

Ferrara, Kathleen, Hans Brunner, and Greg Whittemore. 1991. "Interactive Written Discourse as an Emergent Register." *Written Communication* 8: 8–34.

Fries, Charles Carpenter. 1940. *American English Grammar: The Grammatical Structure of Present-Day American English with Especial Reference to Social Differences or Class Dialects.* New York: Appleton-Century-Crofts.

Ghenu, Mmike. 2005. "Ill Comunication: How Email, Text and Instant Messaging Affect Language." *Varsity* (Univ. of Toronto), Feb. 28, 8.

Grinter, Rebecca E., and Leysia Palen. 2002. "Instant Messaging in Teen Life." In *CSCW 2002: Proceedings of the ACM 2002 Conference on Computer Supported Cooperative Work*, 21–30. New York: Assoc. for Computing Machinery. Available at http://www.cs.colorado.edu/~palen/Papers/grinter-palen-IM.pdf.

Hentschel, Elke. 1998. "Communication on IRC." *Linguistik Online* 1. http://www.linguistik-online.de/irc.htm.

Herring, Susan C. 2003. "Gender and Power in On-line Communication." In *The Handbook of Language and Gender*, ed. Janet Holmes and Miriam Meyerhoff, 202–28. Malden, Mass.: Blackwell.

———. 2004. "Slouching toward the Ordinary: Current Trends in Computer-Mediated Communication." *New Media and Society* 6: 26–36.

Hundt, Marianne, and Christian Mair. 1999. "'Agile' and 'Uptight' Genres: The Corpus-based Approach to Language Change in Progress." *International Journal of Corpus Linguistics* 4: 221–42.

Ito, Rika, and Sali Tagliamonte. 2003. "*Well* Weird, *Right* Dodgy, *Very* Strange, *Really* Cool: Layering and Recycling in English Intensifiers." *Language in Society* 32: 257–79.

Labov, William. 1972. *Sociolinguistic Patterns.* Philadelphia: Univ. of Pennsylvania Press.

Lee, Jennifer 8. 2002. "I Think, Therefore IM." *New York Times*, Sept. 19, G1. Available at http://www.nytimes.com/2002/09/19/technology/circuits/19MESS.html. Alternatively titled "Nu Shortcuts in School R 2 Much 4 Teachers."

Leech, Geoffrey, and Nicholas Smith. 2005. "Extending the Possibilities of Corpus-based Research on English in the Twentient Century: A Prequel to LOB and FLOB." *ICAME Journal* 29: 83–98.

Lenhart, Amanda, Lee Rainie, and Oliver Lewis. 2001. "Teenage Life Online: The Rise of the Instant-Message Generation and the Internet's Impact on Friendships and Family Relationships." Pew Internet and American Life Project. http://www.pewinternet.org/pdfs/PIP_Teens_Report.pdf (accessed Nov. 25, 2004).

Ling, Rich, and Naomi S. Baron. 2007. "Text Messaging and IM: Linguistic Comparison of American College Data." *Journal of Language and Social Psychology* 26: 291–98.

Mair, Christian. 1998. "Corpora and the Study of the Major Varieties of English: Issues and Results." In *The Major Varieties of English: Papers from MAVEN 97*, ed. Hans Lindquist, Staffan Klintborg, Magnus Levin, and Maria Estling, 139–57. Växjö, Sweden: Växjö Acta Wexionensia.

Nardi, Bonnie A., Steve Whittaker, and Erin Bradner. 2000. "Interaction and Outeraction: Instant Messaging in Action." In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, 79–88. New York: Association for Computing Machinery. http://www.acm.org/pubs/contents/proceedings/cscw/358916/.

O'Connor, Amanda. 2005. "Instant Messaging: Friend or Foe of Student Writing." *New Horizons for Learning Online Journal* 11.2. http://www.newhorizons.org/strategies/literacy/oconnor.htm.

Palfreyman, David, and Muhamed al Khalil. 2003. "'A Funky Language for Teenzz to Use': Representing Gulf Arabic in Instant Messaging." *Journal of Computer-Mediated Communication* 9.1. http://jcmc.indiana.edu/vol9/issue1/palfreyman.html.

Paolillo, John. 1999. "The Virtual Speech Community: Social Networks and Language Variation on IRC." *Journal of Computer-Mediated Communication* 4.4. http://jcmc.indiana.edu/vol4/issue4/paolillo.html.

Peters, Hans. 1994. "Degree Adverbs in Early Modern English." In *Studies in Early Modern English*, ed. Dieter Kastovsky, 269–88. Berlin: Mouton de Gruyter.

Poplack, Shana, and Sali Tagliamonte. 1999. "The Grammaticalization of *going to* in (African American) English." *Language Variation and Change* 11: 315–42.

———. 2001. *African American English in the Diaspora*. Malden, Mass.: Blackwell.

Rand, David W., and Tatiana Patera. 1991. *Concorder: Concordance Software for the Macintosh*. Montreal: Les Publications CRM.

Randall, Neil. 2002. "Lingo Online: A Report of the Language of the Keyboard Generation." Unpublished MS. Available at http://arts.uwaterloo.ca/~nrandall/LingoOnline-finalreport.pdf.

Sankoff, D., S. A. Tagliamonte, and E. Smith. 2005. "Goldvarb X: A Multivariate Analysis Application." Dept. of Linguistics, Univ. of Toronto. http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.

Sankoff, Gillian. 1980. *The Social Life of Language*. Philadelphia: Univ. of Pennsylvania Press.

Schiano, Diane J., Coreena P. Chen, Jeremy Ginsberg, Unnur Gretarsdottir, Megan Huddleston, and Ellen Isaacs. 2002. "Teen Use of Messaging Media." *CHI '02: Proceedings of ACM Conference on Human Factors in Computing Systems*, 594–95. New York: Assoc. for Computing Machinery. Available at http://portal.acm.org/ft_gateway.cfm?id=506500&type=pdf&coll=GUIDE&dl=GUIDE&CFID=49265163&CFTOKEN=26387788.

Singler, John Victor. 2001. "Why You Can't Do a VARBRUL Study of Quotatives and What Such a Study Can Show Us." *University of Pennsylvania Working Papers in Linguistics* 7.3: 257–78.

Stoffel, Cornelis. 1901. *Intensives and Down-Toners: A Study in English Adverbs*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Tagliamonte, Sali. 2002. "Comparative Sociolinguistics." In *Handbook of Language Variation and Change*, ed. J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes, 729–63. Malden, Mass.: Blackwell.

———. 2003–2006. "Linguistic Changes in Canada Entering the 21st Century." Research Grant. Social Sciences and Humanities Research Council of Canada (SSHRC). #410-2003-0005. http://individual.utoronto.ca/tagliamonte/.

———. 2006. *Analysing Sociolinguistic Variation.* Cambridge: Cambridge Univ. Press.

———. Forthcoming. "*So* Different and *Pretty* Cool: Recycling Intensifiers in Toronto, Canada." In *Intensifiers,* ed. Belén Mendez-Naya. Special issue of *English Language and Linguistics.*

Tagliamonte, Sali, and Alexandra D'Arcy. 2004a. "*He's like, She's like*: The Quotative System in Canadian Youth." *Journal of Sociolinguistics* 8: 493–514.

———. 2004b. "Mom *Said,* and My Daughter*'s Like*: Tracking the Quotative System through the Generations." Paper presented at the 33rd annual conference on New Ways of Analyzing Variation (NWAV 33), Ann Arbor, Mich., Sept. 30–Oct. 3.

———. 2005. "When People Say, 'I Was Like…': The Quotative System in Canadian Youth." In *Pennsylvania Working Papers in Linguistics* 10.2: 257–72.

———. 2007a. "Frequency and Variation in the Community Grammar: Tracking a New Change through the Generations." *Language Variation and Change* 19: 199–217.

———. 2007b. "The Modals of Obligation/Necessity in Canadian Perspective." *English World-Wide* 28: 47–87.

Tagliamonte, Sali, and Rachel Hudson. 1999. "*Be like* et al. beyond America: The Quotative System in British and Canadian Youth." *Journal of Sociolinguistics* 3: 147–72.

Tagliamonte, Sali, and Jennifer Smith. 2006. "Layering, Change and a Twist of Fate: Deontic Modality in Dialects of English." *Diachronica* 23: 341–80.

Thurlow, Crispin. 2002. "Generation Txt? The Sociolinguistics of Young People's Text-Messaging." *Discourse Analysis Online* 1.1. http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003.html.

Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. "Empirical Foundations for a Theory of Language Change." In *Directions for Historical Linguistics: A Symposium,* ed. W. P. Lehmann and Yakov Malkiel, 95–188. Austin: Univ. of Texas Press.

Werry, Christopher C. 1996. "Linguistic and Interactional Features of Internet Relay Chat." In *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives,* ed. Susan C. Herring, 47–63. Amsterdam: Benjamins.

Yates, Simeon J. 1996. "Oral and Written Linguistic Aspects of Computer Conferencing: A Corpus Based Study." In *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives,* ed. Susan C. Herring, 29–46. Amsterdam: Benjamins.